Iterated and anisotropic marked point processes, with a view to the minicolumn hypothesis

Christoffersen, Andreas Dyreborg

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

ITERATED AND ANISOTROPIC MARKED POINT PROCESSES, WITH A VIEW TO THE MINICOLUMN HYPOTHESIS

BY
ANDREAS DYREBORG CHRISTOFFERSEN

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

Iterated and anisotropic marked point processes, with a view to the minicolumn hypothesis

PhD Dissertation

Andreas Dyreborg Christoffersen

Dissertation submitted August 30, 2019

Dissertation submitted: August 30, 2019

PhD supervisor: Prof. Jesper Møller
Aalborg University

Assistant PhD supervisor: Prof. Rasmus Plenge Waagepetersen
Aalborg University

PhD committee: Associate Professor Poul Svante Eriksen(chairman)
Aalborg University

Doc. Tomás Mrkvicka
University of South Bohemia

Associate Professor Ute Hahn
Aarhus University

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Mathematical Sciences

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-487-4

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2019

Abstract

Spatial statistics is a large branch of statistics, which includes the theory of point processes. A point process is a random and countable collection of points from some space, and its realisations are called point patterns. This subject is of interest in many scientific fields such as ecology, stereology, neurology, and astronomy, where information is collected from certain locations and where these locations, themselves, are of interest. Often, the locations constitute the points of the point patterns, and the information, collected at the locations, is attributed to the points and are referred to as marks.

This thesis deals with point processes and has a twofold focus.

The first focus is on analysing two point patterns each concerning the location and orientations of neurons contained in a section of the cerebral cortex of a human brain. The structuring of neurons is of great interest in neuroscience, since deviations from a "normal" neural structure can be linked to different neurological diseases. One such structure is the so-called minicolumn structure, whose existence is still debatable. This minicolumn structure is of particular interest in the present thesis. The two point patterns consist of three-dimensional coordinates for the locations together with spherical coordinates for the orientations of so-called pyramidal cells – specific cells that are pyramid shaped and thus have a natural orientation. Hence, the data are marked point patterns with orientations being the information collected at the location of the neurons (the opposite is equally true, but seems less appealing). These marked point patterns may simply be viewed as point patterns consisting of points from the product space $\mathbb{R}^3 \times \mathbb{S}^2$. To compare such a point pattern to a proposed model we extend the concept of the so-called K-function (a particular function that summarises certain characteristics of a point pattern) to the product space $\mathbb{R}^d \times \mathbb{S}^k$. Additionally, we apply this function to test for dependence between the spatial (locations) and spherical (orientations) component of the point patterns. As a result we find that the two components are independent and consequently can be modelled separately. Hence, we find and discuss a suitable model for the orientations and a satisfactory model for the locations.

The second focus of this thesis is on some theoretical details on a certain

class of point process models, namely the so-called iterated cluster point processes. Iterated cluster point processes are briefly defined as a set of points, each of which generate a set of offspring points, which in turn generate offspring and so on. These type of point processes are of interest in relation to e.g. population genetics and community ecology, since they naturally describe a reproduction mechanism. We focus on finding a closed term expression for a particular summary function and show, under certain conditions, that the point process converges in distribution as the number of iterations tends to infinity.

The thesis is divided into two parts: the first part of the thesis gives an introduction to the basic concept that are necessary for grasping the concepts that are presented in the second part of the thesis; the second part of the thesis consist of a collection of papers that deals with the aforementioned problems.

Resumé

Rumlig statistik er en gren inden for statistik, der inkluderer teorien om punktprocesser. En punktproces er en stokastisk og tællelig samling af punkter fra et vilkårligt rum, og dens realiseringer kaldes for punktmønstre. Dette emne er af interesse inden for mange videnskabelige områder såsom økologi, stereologi, neurologi og astronomi, hvor information indsamles fra bestemte lokationer, og hvor disse lokationer i sig selv er af interesse. Ofte udgør lokationerne punkterne i punktmønstret, og den yderligere information, indsamlet på lokationerne, henføres til punkterne og bliver omtalt som mærker.

Denne afhandling omhandler punktprocesser og har et todelt fokus.

Det første fokus er på at analysere to punktmønstre, der hver vedrører lokationen og orienteringen af neuroner i et område af hjernebarken fra en menneskehjerne. Organiseringen af neuroner er af stor interesse i neurologi, da denne kan knyttes til forskellige neurologiske sygdomme. Den såkaldte minisøjle-struktur, hvis eksistens stadig diskuteres, er en del af denne organisring. Minisøjle-strukturen er af særlig interesse i denne afhandling. De to punktmønstre består af tredimensionelle koordinater for lokationerne sammen med sfæriske koordinater for orienteringerne af såkaldte pyramideceller – specifikke celler, der er pyramideformet og som derfor har en naturlig orientering. Punktmønstrene består altså af punkter fra produktrummet $\mathbb{R}^3 \times \mathbb{S}^2$. For at sammenligne et sådant punktmønster med en model udvider vi den såkaldte K -funktion (en bestemt funktion, der opsummerer visse egenskaber ved et punktmønster) til produktrummet $\mathbb{R}^d \times \mathbb{S}^k$. Yderligere, anvender vi denne funktion til at teste for afhængighed mellem den rumlige (lokationerne) og sfæriske (orienteringerne) komponent i punktmønstrene. Vi finder, at de to komponenter er uafhængige og derfor kan modelleres separat. Derfor finder vi og diskuterer en passende model til orienteringerne samt en tilfredsstillende model for placeringerne.

Det andet fokus i afhandlingen er på nogle teoretiske detaljer om en bestemt klasse af punktprocesmodeller, nemlig de såkaldte itererede klyngepunktprocesser. Itererede klyngepunktprocesser kan kort defineres som et sæt af punkter, som hver genererer et sæt afkomspunkter, som igen genererer afkom og så fremdeles. Denne type punktprocesser er af interesse i forhold

til f.eks. populationsgenetik og samfundsøkologi, da de naturligt beskriver en reproduktionsmekanisme. Vi fokuserer på at finde et lukket udtryk for en bestemt opsummerende funktion og viser, under visse betingelser, at punktprocessen konvergerer i fordeling, når antal iterationer går mod uendelig.

Afhandlingen er opdelt i to dele: første del af afhandlingen giver en introduktion til de grundlæggende koncepter, der er nødvendigt for at forstå koncepterne præsenteret i anden del af afhandlingen; anden del af afhandlingen består af en samling artikler, der beskæftiger sig med de ovenfornævnte problemer.

Contents

Preface

This thesis is the result of the work done by my co-authors and me during my time as a PhD-student at the Department of Mathematical Sciences at Aalborg University, Denmark. The work is supported by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by the Villum Foundation.

The original objective of this thesis was to investigate the organisation of pyramidal cells in the human cerebral cortex, with a particular focus on the so-called minicolumn hypothesis. This is of great interest for biologists since it is suspected that different minicolumnar structures are related to different neurological disorders. In the process of finding a suitable point process model for to describe the neural structure a lot of time was spent on theoretical results for so-called iterated cluster point processes, which now constitutes a great part of this thesis.

The thesis is structured in two parts: part I gives a introduction to the minicolumn hypothesis, the data considered in this thesis, which are subsets of the product space $\mathbb{R}^3 \times \mathbb{S}^2$, as well as a brief introduction to the most relevant theory of spatial point processes; part II is a collection of the three papers:

A: Møller, J., Christensen, H. S., Cuevas-Pacheco, F., and Christoffersen, A. D. (2019). Structured space sphere point processes and K-functions. To appear in *Methodology and Computing in Applied Probability.* Available at `https://doi.org/10.1007/s11009-019-09712-w`.

B: Møller, J. and Christoffersen, A. D. (2018). Pair correlation functions and limiting distributions of iterated cluster point processes. *Journal of Applied Probability,* 55:789–809.

A: Christoffersen, A. D., Møller, J., and Christensen, H. S. (2019). Modelling columnarity of pyramidal cells in the human cerebral cortex. Submitted to *Journal of the Royal Statistical Society: Series C (Applied Statistics).* Available on arXiv:1908.05065.

In the first paper we investigate whether there is a dependence between the two components (that is the points of $\mathbb{R}^3$ and $\mathbb{S}^2$, respectively) of the datasets,

and discuss a suitable model for the spherical component (that is the points in $\mathbb{S}^2$). Theoretical details on iterated cluster point processes are discussed in the second paper and finally in the third paper a model for columnar structures, which makes use of a special case of the iterated cluster point process model, is presented. It should be noted that the notation may be inconsistent over the three papers and background knowledge may be presented multiple times.

I would like to extend my sincerest gratitude to my supervisor Professor Jesper Møller for sharing his extensive knowledge and for his always honest opinions. His dedication and industriousness is truly inspiring. I would also like to thank Associate Professor Ege Rubak for his availability and kindness whenever any software caused me trouble as well as Professor Jens R. Nyengaard and his team for supplying the data, and for helpful discussions on the biological aspect of the thesis. Finally, I would like to thank my loving and caring partner Heidi S. Christensen, not only for her collaboration and many fruitful professional discussions through the last three years, but also for her love and moral support and for sticking around through the difficult periods.

Andreas D. Christoffersen
Aalborg University, August 30, 2019

Part I

Background

Background

1 Introduction

In many areas of science, such as ecology, stereology, neurology, and astronomy, datasets consisting of the locations of events (called points) arise. Such datasets are referred to as point patterns, and often the interest is to investigate the underlying structure of the points. A typical example of a point pattern is the locations of trees in a forest or sub-region of a forest. In such a case relevant questions could be how likely it is to observe two trees in close proximity and possibly how this probability is affected by the presence/absence of different nutrients in the soil or characteristics of the trees, such as height or species. For many applications additional information, such as the species of the tree, are recorded and attributed to each point. This information is called a mark and can be treated in many ways depending on the nature of the mark and the application. As an extension to the tree example the marks could be the generations of the trees (of course this assignment of generations would begin at an arbitrary time, since some of the eldest trees will have no observable parents). With that additional information one could investigate the reproduction mechanism of the trees, that is how are the offspring trees scattered from the parent tree; or basically measuring how far the apple fall from the tree. In that case the points would be contained in $\mathbb{R}^2$ and the marks in $\mathbb{N}$.

Obviously, more complicated types of point patterns arise in practise and we consider such an instance in Section 2, where the open problem of the minicolumn hypothesis along with two associated point patterns are presented. Following, in Section 3 a brief presentation on the basic, and for this thesis essential, concepts for spatial point process. Finally, Section 4 gives some closing remarks and summarise the outcome of each of the three papers contained in the second part of this thesis.

2 The minicolumn hypothesis

Neurons are the fundamental cells of the nervous system that receive, process and transmit signals/information and are in essence responsible for the actions of its host (humans or animals), such as feeling, seeing, thinking, and moving. A neuron is a cell that typically consists of a soma, multiple dendrites, and a single axon. The soma is also called the cell body and contains a nucleus, which in layman terms is referred to as the control centre of the cell, and contains the genetic material of the cell as well as the nucleolus. The soma varies in size and shape depending on the type of neuron and for instance the pyramidal cells, which makes up approximately 75 % to 80 % of all neurons (Buxhoeveden and Casanova, 2002), have pyramid-shaped somas. The soma processes information received by the dendrites and passes the processed information on to other neurons via the axon. Thus, the dendrites and the axon constitute the links connecting neurons. Understanding how neurons are connected and organised is key in understanding the workings of the nervous system, including the brain.

The neocortex is part of the cerebral cortex and is organised in six layers (layers I-VI) each with its own unique neural structure. This layering structure is part of the horizontal (relative to the so-called pial surface of the brain) organisation of the cortex. Correspondingly, the neurons are vertically organised and in particular the neurons of the cortical layers II-VI are hypothesised to be organised in small linear aggregates which Lorente de Nó (1938) suggested to be functional elementary units. Later Mountcastle (1978) further defined these linear aggregates as minicolumns which constitute the smallest level of vertical organisation and are, in humans, estimated to have a diameter of 35 µm to 60 µm and consist of 80 neurons to 100 neurons. The structure and functionality of minicolumns is still being debated today. A brief discussion on this is found in Rafati et al. (2016). Although, or maybe because, minicolumnarity in the neocortex is still debated, it is important to understand its structure and functionality not at least because the minicolumn structure may be associated to different psychological and neurological diseases, such as schizophrenia (Casanova, 2007), Alzheimer's disease (Esiri and Chance, 2006), autism (Casanova et al., 2006), and Down's syndrome (Buxhoeveden and Casanova, 2002).

2.1 State of the art

The minicolumn hypothesis has been extensively studied in the biological literature (see Buxhoeveden and Casanova (2002) or more recently Rafati et al. (2016) and references therein), but only few statistical analyses have been performed (see Skoglund et al., 2004; Cruz et al., 2005, 2008; Rafati et al., 2016). Specifically, Skoglund et al. (2004) analysed stacks of two-dimensional im-

ages imitating the three-dimensional organisation of neurons in mice brains. Cruz et al. (2005) applied a two-dimensional density map method in order to quantify the average minicolumn, using data consisting of two-dimensional coordinates of cell centres in two thin tissue samples from the brain of a monkey. Cruz et al. (2008) extended the concept of the two-dimensional density map to create a theoretical model that can construct three-dimensional representations of neurons based on two-dimensional coordinates. These three studies are based on two-dimensional data, which means that the results only approximate the reality. In contrast to these studies Rafati et al. (2016) considered three-dimensional point pattern data describing the locations of pyramidal cells and found in three out of four datasets some degree of columnarity. Though a model was proposed in this paper it was not fitted to data.

2.2 The datasets

The focus of this thesis was to analyse two datasets similar to those of Rafati et al. (2016), that is three-dimensional locations of pyramidal cells' nucleolus centre, but with the additional information of the orientations of the cells. A pyramidal cell does as mentioned have a pyramid-shaped soma, and at the top of the pyramid the apex, which is the base of the so-called apical dendrite, is found. The orientation of each of the pyramidal cells was measured as the location of the apex relative to the cells' nucleolus centre. These orientations can be thought of as marks, yielding marked point patterns with points restricted to $W \subset \mathbb{R}^3$ and marks in $\mathbb{S}^2$, where W is the observation region of the points and $\mathbb{S}^2$ is the unit sphere. Alternatively, the data may be considered point patterns in the product space $W \times \mathbb{S}^2$. The two datasets were collected from regions of layer III and V of the fourth Brodmann area of the human brain, which is the primary motor cortex, and we refer to the data as `L3` and `L5`, respectively. The observation regions for `L3` and `L5` consist of 634 and 548 points and are of dimension $[0, 492.7] \times [0, 132.03] \times [0, 407.7]$µm and $[0, 488.44] \times [0, 138.33] \times [0, 495.4]$µm, respectively. Finally the locations are collected in such a way that the z-axis is perpendicular to the pial surface of the brain, which implies that the minicolumns – if they exist – are parallel to the z-axis.

3 Point processes

A point process, X, is the mathematical object used for modelling point patterns. It is a random countable subset of some space S, that is often required to be locally finite; that is, $n(X \cap B) < \infty$ for any bounded $B \subseteq S$, where $n(\cdot)$ denotes the cardinality. For a more rigorous definition see Daley and Vere-Jones (2003). For most applications $S \subseteq \mathbb{R}^d$, but point patterns on other

metric spaces, such as $\mathbb{S}^k = \left\{ \xi \in \mathbb{R}^{k+1} : \|\xi\| = 1 \right\}$ or $\mathbb{R}^d \times \mathbb{S}^k$ may arise.

A point process, X, on $\mathbb{R}^d$ is said to be stationary if the distribution of X is invariant under translation and isotropic if its distribution is invariant under rotation about the origin. If X is defined on $\mathbb{S}^k$ stationarity is not defined, but X is said to be isotropic if its distribution is invariant under rotation on $\mathbb{S}^k$ (corresponds to rotating around the origin).

3.1 Functional summaries

Functional summaries are functions that reflect certain characteristics of a model and can be empirically estimated for the data. These functions are useful for fitting models as well as model control.

The n'th order moment measures

The n'th order intensity function $\rho^{(n)} : S^n \to [0, \infty)$, $n \in \mathbb{N} := \{1, 2, \ldots\}$, can heuristically be interpreted as the simultaneous probability of observing a point in each of the infinitesimal regions of size $\mathrm{d}\xi_1, \ldots, \mathrm{d}\xi_n$ and containing $\xi_1, \ldots, \xi_n$, respectively. Given that $\rho^{(n)}$ exists, it is defined such that is satisfy

$$\begin{aligned} &\mathrm{E} \sum_{\xi_1, \ldots, \xi_n \in X}^{\neq} \mathbb{I}\left[(\xi_1, \ldots, \xi_n) \in C\right] \\ &\quad = \int \cdots \int \mathbb{I}\left[(\xi_1, \ldots, \xi_n) \in C\right] \rho^{(n)}(\xi_1, \ldots, \xi_n) \, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_n, \quad C \subseteq S^n \end{aligned}$$

where $\neq$ above the summation means that the points $\xi_1, \ldots, \xi_n$ are pairwise distinct and $\mathbb{I}(\cdot)$ denotes the indicator function.

The most common functional summary is the first order intensity function $\rho := \rho^{(1)}$ which is simply referred to as the intensity function. A point process with a constant intensity function $\rho(\cdot) \equiv \rho$ is said to be homogeneous.

The ratio of the second and first order product densities (given that they exist)

$$g(\xi_1, \xi_2) = \frac{\rho^{(2)}(\xi_1, \xi_2)}{\rho(\xi_1)\rho(\xi_2)}$$

with $a/0 = 0$ for $a \geq 0$, is called the pair correlation function (PCF). Intuitively, if for some model $g(\xi_1, \xi_2) < 1$ ($g(\xi_1, \xi_2) > 1$) it is unlikely (likely) to jointly observe two points at ξ_1 and ξ_2 as opposed to independently observing a point at each of the two locations. In turn, if $S \subseteq \mathbb{R}^d$ and the PCF is isotropic, that is (when abusing notation) $g(\xi_1, \xi_2) = g(\|\xi_1 - \xi_2\|)$, then $g(r) < 1$ ($g(r) > 1$) can be interpreted as repulsion (clustering) of the points at distance r. Note that, for $S \not\subseteq \mathbb{R}^d$ a different distance metric needs to be imposed, and the particular case of $S \subseteq \mathbb{R}^d \times \mathbb{S}^k$ is discussed in Paper A.

K-functions

For a point process, X, for which the PCF is invariant under translation, that is, (again abusing notation) $g(\xi_1, \xi_2) = g(\xi_1 - \xi_2)$, we may define a different class of second order functional summaries called the second order reduced moment measure, $\mathcal{K}$. This measure is defined as

$$\mathcal{K}(B) = \int_B g(\xi)\,\mathrm{d}\xi, \quad B \subseteq S. \tag{1}$$

If X is stationary, $\rho\mathcal{K}(B)$ can be interpreted as the expected number of further points in B given that X has a point at the origin. This interpretation comes from the reduced Palm distribution, which will not be discussed further but the definition see e.g. Coeurjolly et al. (2017) or Appendix C of Møller and Waagepetersen (2004).

For $S \subseteq \mathbb{R}^d$ the most common choice of structuring element, B in (1), is the ball with radius r and centre at the origin, denoted $b(0, r)$. This choice of B yields the K-function, $K(r) = \mathcal{K}(b(0, r))$, which was originally introduced for homogeneous point processes by Ripley (1976). The definition was later extended to the inhomogeneous case by Baddeley et al. (2000). This particular choice of structuring element is rotationally symmetric and therefore of little use when the goal is to detect columnar structures, which are inherently anisotropic i.e. not rotationally symmetric. For this, Møller et al. (2016) introduced the cylindrical K-function, $K_u(r, t) = \mathcal{K}(\mathcal{O}_u C(r, t))$, where

$$C(r, t) = \left\{(x_1, \ldots, x_d) \in \mathbb{R}^d \,:\, \|(x_1, \ldots, x_{d-1})\| \leq r, |x_d| \leq t\right\}$$

is the cylinder with height $2t$, base radius r, centre of mass at the origin, and direction $e_d = (0, \ldots, 0, 1) \in \mathbb{R}^d$, and $\mathcal{O}_u$ is the rotation operation satisfying $\mathcal{O}_u e_d = u$ where $u \in \mathbb{R}^d$ is arbitrary with $\|u\| = 1$.

For $S \subseteq \mathbb{R}^3$, Redenbach et al. (2009) consider the case when B is the double spherical cone, which is also useful for detecting anisotropy. However, for our application it seems more natural to consider the cylindrical K-function, since we are specifically modelling cylindrical structures.

In Paper A we define the natural extension of the inhomogeneous K-function to space-sphere point processes, i.e. when $S \subseteq \mathbb{R}^d \times \mathbb{S}^k$. Further, we discuss a certain separability property of the first and second order intensity functions and its implications for the space-sphere K-function. Particularly, with the datasets `L3` and `L5` in mind we use the space-sphere K-function to test for independence between locations and orientations. Finally, we perform a simulation study to clarify how well the space-sphere K-function detects non-separability.

Other functional summaries

The empty space function, F, the nearest-neighbour distribution function, G, and the J-function are other isotropic functional summaries that are commonly considered. For a stationary point process, $X \subset \mathbb{R}^d$, these functional summaries are defined as

$$\begin{aligned} F(r) &= P(X \cap b(0,r) \neq \emptyset), \\ G(r) &= \frac{1}{\rho|B|}\mathrm{E} \sum_{\xi \in X \cap B} \mathbb{I}\left[(X \backslash \{\xi\}) \cap b(\xi, r) \neq \emptyset\right], \\ J(r) &= \frac{1 - G(r)}{1 - F(r)} \qquad \text{for } F(r) < 1. \end{aligned}$$

Clearly, the F-function is the probability of observing a point within distance r of the origin (or any other point of S, since X is required to be stationary). The G-function can be interpreted as the probability of observing another point within distance r of the typical point (again this interpretation comes from the Palm distribution, see Coeurjolly et al., 2017; Møller and Waagepetersen, 2004). In turn, the J-function is interpreted as the probability of not observing a point within distance r of a typical point relative to the probability of not observing a point within distance r of the origin. Thus $J(r) < 1$ ($J(r) > 1$) indicates clustering (repulsion) at scale r.

Model control

To investigate how well a fitted model describe a dataset we apply the so-called global rank envelope test (introduced by Myllymäki et al., 2017), which is a procedure that first approximates the distribution of a functional summary under some null model and then compares the empirical functional summary of the dataset to this approximated distribution. This procedure yields a p-interval and the results can be graphically displayed, which is a huge advantage that allows the user to see more details on how the dataset deviates from the null model. Further, a refinement of this method is introduced in Mrkvička et al. (2018), which requires less simulations and yields a single p-value.

3.2 Binomial and Poisson point processes

The binomial point process is a very simple model, where the number of points is fixed. Specifically, X is said to be a binomial point process, on $B \subseteq S$, if n (a fixed number) points are independent and identically distributed with some density f, on B. This model is important in relation to the definition of the so-called Poisson line cluster point process, which is described later.

There we consider the simplest binomial points process with $f(\xi) = 1/|B|$, that is, the points of X are uniformly scattered in B.

The most important point process model is the Poisson point process. It is fully specified by its intensity function, ρ, (implying that there is no interaction between points) and is easily defined through the binomial point process as follows: X is a Poisson point process on S with intensity function ρ if, for any bounded set $B \subseteq S$, $n(X \cap B)$ is Poisson distributed with rate $\mathrm{E}[n(X \cap B)] = \int_B \rho(\xi)\,\mathrm{d}\xi$ and conditioned on $n(X \cap B) = n$ for any $n \in \mathbb{N}$, $X \cap B$ is a binomial point process with n points and density $f(\xi) = \rho(\xi) / \int_B \rho(\xi)\,\mathrm{d}\xi$.

A homogeneous Poisson point process is also referred to as complete spatial randomness (CSR) and is often used as a reference model, since it has no underlying structure. How the observed point pattern deviates from CSR can be very useful information in regards to understanding the data and in turn finding a suitable model.

In Paper A we consider an inhomogeneous Poisson point process model for the orientations of the pyramidal cell data, where the intensity function is a mixture of the so-called Kent and Watson distributions. The Kent distribution is briefly a generalisation of the anisotropic Gaussian distribution to $\mathbb{S}^2$, while the Watson distribution (in our case) generate points that are uniformly distributed on a great circle and then displaced away from this great circle by a von-Mises distribution; the von-Mises distribution is a generalisation of the Gaussian distribution to a circle. The fitted models are found to be suitable.

3.3 Iterated cluster point processes

Paper B deals with so-called iterated cluster point processes which are basically discrete time Markov chains of point processes, $G_0, G_1, \ldots \subset \mathbb{R}^d$, where the i'th state briefly can be defined as follows. Let the state G_{i-1} for $i \in \mathbb{N}$ be a given point process on S, and associate to each point $\xi \in G_{i-1}$ a point process, X_ξ. Then the i'th state is given by the superposition $G_i = \bigcup_{\xi \in G_{i-1}} (X_\xi + \xi)$, where $X_\xi + \xi$ denotes the translation of X_ξ by ξ and where the initial state, G_0, is an arbitrary point process. The particular model is properly introduced in Paper B.

This type of model is interesting in many contexts and is e.g. studied in relation to population genetics and community ecology, tree and protein locations with different scales of clustering (see Shimatani, 2010; Wiegand et al., 2007; Andersen et al., 2018). In Paper B we extend and elaborate the results obtained by Shimatani (2010), which discusses a certain iterated cluster point process that itself extends the work by Malécot which is discussed in Felsenstein (1975). The iterated cluster point process of Shimatani (2010), further, allow for independent migration, which means that the i'th state is specified as $G_i = \bigcup_{\xi \in G_{i-1}} (X_\xi + \xi) \bigcup Z_i$, where $Z_1, Z_2, \ldots$ are independent point pro-

cesses and Z_i is independent of $G_0, \ldots, G_{i-1}$. In Shimatani (2010) it is further assumed that G_0 is a Poisson point process, though it is mentions that the PCF of G_i is tractable as long as the the PCF of G_0 can be written as a certain convolution (details are included in Paper B). However, it is not argued that such a point process exists, except for the Poisson point process. Finally, Shimatani (2010) shows that under certain conditions the PCF of G_i, denoted g_i, converges as $i \to \infty$, but not that the Markov chain has a time stationary distribution, i.e. that the Markov chain converges in distribution, and that the point process associated to this time stationary distribution has the PCF, g_i. These arguments are given in Paper B under more general conditions and for an extended model.

The so-called (modified) Thomas point process (Thomas, 1949) is a special case of the iterated cluster point process and becomes important later when the Poisson line cluster point process is to be defined. Specifically, G_1 is said to be a Thomas point process if G_0 is a stationary Poisson point process, Z_1 is empty, and X_ξ for $\xi \in G_0$ are independent Poisson point processes with intensity

$$\rho_{X_\xi}(\eta) = \frac{\alpha}{2\pi\sigma^2} \exp\left\{ \frac{-\|\eta\|^2}{2\sigma^2} \right\} \qquad \text{for } \eta \in S.$$

Note that ρ_{X_ξ} is proportional to the density of the isotropic and zero-mean Gaussian distribution, $N_d(0, \sigma^2 I_d)$, where I_d is the $d \times d$ identity matrix and the proportionality constant $\alpha > 0$ is the expected number of points in X_ξ.

3.4 Determinantal point processes

In Paper B we discuss choices of G_0 which ensure that the PCF of G_i for $i \in \mathbb{N}$ is tractable, and two such point processes are the so-called determinantal and permanental point processes. The determinantal point process (DPP) is defined in terms of its $n'th$ order intensity functions as follows. Let $\mathbb{C}$ denotes the complex plane, and consider a function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, which we require to be non-negative. Then a point process, X, with n'th order intensity function $\rho^{(n)}$ is called a DPP with kernel C, if

$$\rho^{(n)}(\xi_1, \ldots, \xi_n) = \det[C](\xi_1, \ldots, \xi_n) \qquad \text{for } \xi_1, \ldots, \xi_n \in \mathbb{R}^d, \tag{2}$$

where $\det[C](\xi_1, \ldots, \xi_n)$ is the determinant of the $n \times n$ matrix with entry (i, j) given by $C(\xi_i, \xi_j)$, and $n \in \mathbb{N}$. To obtain the definition for permanental point processe one can simply replace the determinant with the permanent in the above definition. For more thorough and detailed expositions on DPPs see Lavancier et al. (2015, 2014) and for permanental point processes see McCullagh and Møller (2006). Note that a point process is not, in general, uniquely characterised by its moments. However, it can be shown that, for

any kernel C, there only exists one point process that satisfy (2) (Lemma 4.2.6 in Hough et al., 2009).

The Poisson point process corresponds to a DPP in the special case where $C(\xi_1, \xi_2) = 0$ whenever $\xi_1 \neq \xi_2$ and satisfies $\rho^{(n)}(\xi_1, \ldots, \xi_n) = \rho(\xi_1) \cdots \rho(\xi_n)$ for any $n \in \mathbb{N}$. If C is Hermitian, $\rho^{(n)}(\xi_1, \ldots, \xi_n) \leq \rho(\xi_1) \cdots \rho(\xi_n)$ (the opposite is true for permanental point processes). This implies that DPPS are repulsive, and in turn $g(\xi_1, \xi_2) \leq 1$

The case where G_1 of the iterated cluster point process is defined as the Thomas point process, but with the exception that the initial generation, G_0, is a DPP is considered and applied to data in Paper C. This model will be referred to as the determinantal Thomas point process.

3.5 Poisson line cluster point processes

The Poisson line cluster point process (PLCPP) was presented in the same paper as the cylindrical K-function (Møller et al., 2016), and is to the best of our knowledge the only model for columnar structures. Hence, we find it natural to apply this model to the pyramidal cell data, `L3` and `L5`. According to the minicolumn hypothesis, the minicolumns are expected to be perpendicular to the pial surface of the brain, and in turn this implies (at least for small sample regions) that the minicolumns are approximately parallel. Therefore, we restrict attention to the degenerate PLCPP, which is a hierarchical model defined as follows. Let $X_{xy} \subset \mathbb{R}^2$ be a Thomas point process in the (x, y)-plane with centre process Φ. Then $X = \{(x, y, z) : (x, y) \in X_{xy}, z \sim Unif(0, a)\}$, with $a < \infty$, is a degenerate PLCPP on $S = \{(x, y, z) \in \mathbb{R}^3 : z \in [0, a]\}$. Note that for X_{xy} restricted to a bounded region $B_{xy} \subset \mathbb{R}^2$, $X_z = \{z : (x, y, z) \in X \cap (B_{xy} \times [0, a])\}$ is a binomial point process on $[0, a]$ consisting of $n(X_{xy} \cap B_{xy})$ points and with a uniform density. It should be clear from the definition that the degenerated PLCPP consists of cylindrical clusters of points centred around the lines $l_i = \{(\xi_i, \eta_i, z) : z \in [0, a]\}$ for $(\xi_i, \eta_i) \in \Phi$, which are parallel to the z-axis. The prefix 'degenerate' refers to the fact that these lines are parallel. Under the general definition of the PLCPP in Møller et al. (2016), X need not be degenerate and it is defined on all of $\mathbb{R}^3$. However, these assumptions are convenient for the application in Paper C.

In Paper C we generalise the degenerate PLCPP in two ways: we let X_{xy} be a determinantal Thomas point process; and conditioned on $X_{xy} = x_{xy}$ we let X_z be a Markov random field indexed by the points of x_{xy}. Hence, to each fixed location in x_{xy} we associate a random variable following a distribution which depends only on other points within some neighbouring region. By this construction we allow for aggregation, repulsion, or even both.

4 Overview

In summary, this thesis consists of the three papers in part II which briefly contains the following scientific contributions.

In Paper A we extend the K-function to the product space $\mathbb{R}^d \times \mathbb{S}^k$ and discuss separability of the spatial and spherical components (cell locations and orientations). We apply this K-function to two datasets and perform a simulation study investigating how well this K-function detects non-separability. Additionally, a model for the spherical component (cell orientationes) of the pyramidal cell point patterns, is discussed and fitted to the data. Furthermore, we find independence between the spatial and spherical components, using this newly introduced K-function, implying that the two components may be modelled separately.

In Paper B an expression for the PCF of iterated cluster point processes is given, when the PCF of the initial generation and of the noise point processes are of a certain form. We discuss for which point processes the PCF has this form, which, as previously mentioned, includes the determinantal and permanental point processes. It is also shown that under reasonable conditions the distribution of iterated cluster point processes converges as the number of generations tends to infinity and we give the expression of the PCF of the point process following this limiting distribution.

In Paper C we find a model for the spatial component (nucleolus locations) of the pyramidal cell data, which in two ways generalise the Poisson line cluster point proecss: first by replacing the Thomas point process in the plane by the determinantal Thomas point process; second by replacing the uniformly distributed z-coordinate by a Markov random field indexed by the (x, y)-coordinates that we condition on.

References

Andersen, I. T., Hahn, U., Christensen, E. A., Nejsum, L. N., and Jensen, E. B. V. (2018). Double Cox cluster processes - with applications to photoactivated localization microscopy. *Spatial Statistics*, 27:58–73.

Baddeley, A., Møller, J., and Waagepetersen, R. P. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350.

Buxhoeveden, D. P. and Casanova, M. F. (2002). The minicolumn hypothesis in neuroscience. *Brain*, 125:935–951.

Casanova, M. F. (2007). Schizophrenia seen as a deficit in the modulation of cortical minicolumns by monoaminergic systems. *International Review of Psychiatry*, 19:361–372.

Casanova, M. F., van Kooten, I. A. J., Switala, A. E., van Engeland, H., Heinsen, H., Steinbusch, H. W. M., Hof, P. R., Trippe, J., Stone, J., and Schmitz, C. (2006). Minicolumnar abnormalities in autism. *Acta Neuropathologica (Berl)*, 112:287–303.

Coeurjolly, J. F., Møller, J., and Waagepetersen, R. P. (2017). A tutorial on Palm distributions for spatial point processes. *International Statistical Review*, 85:404–420.

Cruz, L., Buldyrev, S. V., Peng, S., Roe, D. L., Urbanc, B., Stanley, H. E., and Rosene, D. L. (2005). A statistically based density map method for identification and quantification of regional differences in microcolumnarity in the monkey brain. *Journal of Neuroscience Methods*, 141:321–332.

Cruz, L., Urbanc, B., Inglis, A., Rosene, D. L., and Stanley, H. E. (2008). Generating a model of the three-dimensional spatial distribution of neurons using density maps. *NeuroImage*, 40:1105–1115.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer-Verlag, New York, second edition.

Esiri, M. M. and Chance, S. A. (2006). Vulnerability to Alzheimer's pathology in neocortex: the roles of plasticity and columnar organization. *Journal of Alzheimer's Disease*, 9:79–89.

Felsenstein, J. (1975). A pain in the torus: Some difficulties with models of isolation by distance. *The American Naturalist*, 109:359–368.

Hough, J. B., Krishnapur, M., Peres, Y., and Viràg, B. (2009). *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. American Mathematical Society, Providence.

Lavancier, F., Møller, J., and Rubak, E. (2014). Determinantal point process models and statistical inference (extended version). available on arXiv:1205.4818.

Lavancier, F., Møller, J., and Rubak, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77:853–877.

Lorente de Nó, R. (1938). The cerebral cortex: Architecture, intracortical connections, motor projections. In Fulton, J. F., editor, *Physiology of the Nervous System*, pages 274–301. Oxford University Press, Oxford.

McCullagh, P. and Møller, J. (2006). The permanental process. *Advances in Applied Probability*, 38:873–888.

Møller, J., Safavimanesh, F., and Rasmussen, J. G. (2016). The cylindrical K-function and Poisson line cluster point process. *Biometrika*, 103:937–954.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, Boca Raton, Florida.

Mountcastle, V. B. (1978). *The Mindful Brain: Cortical Organization and the Group-selective Theory of Higher Brain Function*. MIT Press, Cambridge.

Mrkvička, T., Hahn, U., and Myllymäki, M. (2018). A one-way ANOVA test for functional data with graphical interpretation. Available on arXiv:1612.03608.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

Rafati, A. H., Safavimanesh, F., Dorph-Petersen, K. A., Rasmussen, J. G., Møller, J., and Nyengaard, J. R. (2016). Detection and spatial characterization of minicolumnarity in the human cerebral cortex. *Journal of Microscopy*, 261:115–126.

Redenbach, C., Säkkä, A., Freitag, J., and Schladitz, K. (2009). Anisotropy analysis of pressed point processes. *AStA Advances in Statistical Analysis*, 93:237–261.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13:255–266.

Shimatani, I. K. (2010). Spatially explicit neutral models for population genetics and community ecology: Extensions of the Neyman-Scott clustering process. *Theoretical Population Biology*, 77:32–41.

Skoglund, T. S., Pascher, R., and Berthold, C. H. (2004). Aspects of the organization of neurons and dendritic bundles in primary somatosensory cortex of the rat. *Neuroscience Research*, 50:189–198.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, 36:18–25.

Wiegand, T., Gunatilleke, S., Gumatilleke, N., and Okuda, T. (2007). Analyzing the spatial structures of a Sri Lankan tree species with multiple scales of clustering. *Ecology*, 88:3088–3102.

Part II

Papers

Paper A

Structured space-sphere point processes and K-functions

Jesper Møller, Heidi S. Christensen, Francisco Cuevas-Pacheco,
and Andreas D. Christoffersen

The paper is to appear in
Methodology and Computing in Applied Probability.

The layout has been revised.

Abstract

This paper concerns space-sphere point processes, that is, point processes on the product space of $\mathbb{R}^d$ (the d-dimensional Euclidean space) and $\mathbb{S}^k$ (the k-dimensional sphere). We consider specific classes of models for space-sphere point processes, which are adaptations of existing models for either spherical or spatial point processes. For model checking or fitting, we present the space-sphere K-function which is a natural extension of the inhomogeneous K-function for point processes on $\mathbb{R}^d$ to the case of space-sphere point processes. Under the assumption that the intensity and pair correlation function both have a certain separable structure, the space-sphere K-function is shown to be proportional to the product of the inhomogeneous spatial and spherical K-functions. For the presented space-sphere point process models, we discuss cases where such a separable structure can be obtained. The usefulness of the space-sphere K-function is illustrated for real and simulated datasets with varying dimensions d and k.

1 Introduction

Occasionally point processes arise on more complicated spaces than the usual space $\mathbb{R}^d$, the d-dimensional Euclidean space, as for spatio-temporal point processes, spherical point processes or point processes on networks (see e.g. Dvořák and Prokešová, 2016; Lawrence et al., 2016; Møller and Rubak, 2016; Baddeley et al., 2017, and the references therein for details on such point processes). In this paper we consider *space-sphere point processes* that live on the product space $S = \mathbb{R}^d \times \mathbb{S}^k$, where $\mathbb{S}^k = \{u \in \mathbb{R}^{k+1} : \|u\|_{k+1} = 1\}$ is the k-dimensional unit sphere, $\|\cdot\|_k$ denotes the usual distance in $\mathbb{R}^k$, and $d, k \in \{1, 2, \ldots\}$. For each point $(y, u) \in S$ belonging to a given space-sphere point process, we call y its spatial component and u its spherical component. Assuming local finiteness of a space-sphere point process, the spatial components constitute a locally finite point process in $\mathbb{R}^d$, but the spherical components do not necessarily form a finite point process on $\mathbb{S}^k$. However, in practice the spatial components are only considered within a bounded window $W \subset \mathbb{R}^d$, and the associated spherical components do constitute a finite point process.

One example is the data shown in Figure A.1 that consists of the location and orientation of a number of pyramidal neurons found in a small area of a healthy human's primary motor cortex. More precisely, the locations are three-dimensional coordinates each describing the placement of a pyramidal neuron's nucleolus, and the orientations are unit vectors pointing from a neuron's nucleolus toward its apical dendrite. These data can be considered as a realisation of a space-sphere point process with dimensions $d = 3$ and $k = 2$, where the spatial components describe the nucleolus locations and the

spherical components are the orientations. How neurons (of which around 75 to 80 % are pyramidal neurons) are arranged have been widely discussed in the literature. Specifically, it is hypothesised that neurons are arranged in columns perpendicular to the pial surface of the brain. This hypothesis, referred to as the minicolumn hypothesis, have been studied for more than half a century (see e.g. Lorente de Nó, 1938; Mountcastle, 1978; Buxhoeveden and Casanova, 2002), and it is believed that deviation from such a columnar structure is linked with neurological diseases such as Alzheimers and schizophrenia.

Another example is the time and geographic location of fireballs, which are bright meteors reaching a visual magnitude of -3 or brighter. They are continually recorded by U.S. Government sensors and made available at `http://neo.jpl.nasa.gov/fireballs/`. We can consider fireball events as a space-sphere point process with dimensions $d = 1$ and $k = 2$, where the time and locations are the spatial and spherical components, respectively. Figure A.2 shows the location of fireballs on the globe (identified with the unit sphere) observed over a time period of about 606 weeks.

The paper is organised as follows. In Section 2, we define concepts related to space-sphere point processes and give some natural examples of such processes. In Section 3, we define the space-sphere K-function, a functional summary statistic which is analogue to the space-time K-function when $d = 2$ and $\mathbb{S}^k$ is replaced by the time axis (Diggle et al., 1995; Gabriel and Diggle, 2009; Møller and Ghorbani, 2012). The space-sphere K-function is defined in terms of the pair correlation function which is assumed to have a certain stationary form. In the case where both the intensity and pair correlation function have a specific separable structure discussed in Section 4, the space-sphere K-function is shown to be proportional to the product of the spatial K-function (Baddeley et al., 2000) and the spherical K-function (Lawrence et al., 2016; Møller and Rubak, 2016). Further, an unbiased estimate is given in Section 5. In Section 6, the usefulness of the space-sphere K-function is illustrated for the fireball and neuron data as well as for simulated data, and it is e.g. seen how the K-function may be used to test for independence between the spatial and spherical components.

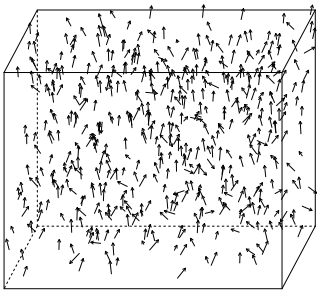

Fig. A.1: Location and orientation of pyramidal neurons in a small section of a human brain. For details, see Section 6.2.

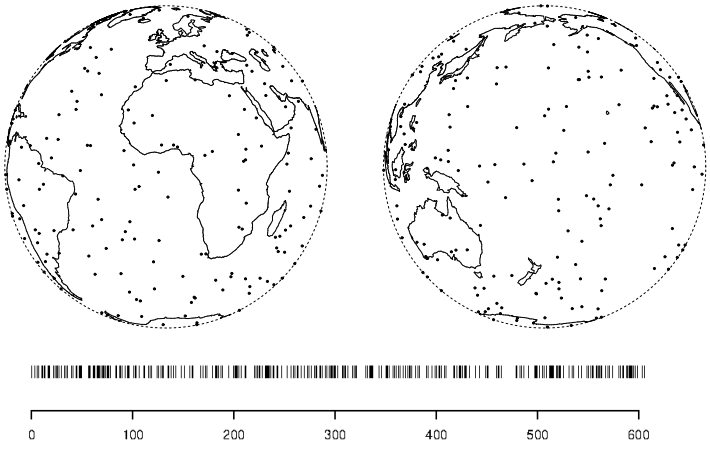

Fig. A.2: Top: orthographic projection of the fireball locations. Bottom: time of fireball events measured in weeks. For details, see Section 6.1.

2 Preliminaries

2.1 Setting

Throughout this paper we consider the following setting.

Equip $\mathbb{R}^d$ with the Lebesgue measure $|A| = \int_A \mathrm{d}y$ and $\mathbb{S}^k$ with Lebesgue/surface measure $\nu(B)$, where $A \subseteq \mathbb{R}^d$ and $B \subseteq \mathbb{S}^k$ are Borel sets. Thus, the product space $S = \mathbb{R}^d \times \mathbb{S}^k$ is equipped with Lebesgue measure μ given by $\mu(A \times B) = |A|\nu(B)$.

Let X be a simple locally finite point process on S, that is, we can view X as a random subset of S such that the restriction $X_B = X \cap B$ of X to any bounded set $B \subset S$ is finite. We call X a *space-sphere point process*, and assume that it has intensity function ρ with respect to μ and pair correlation function g with respect to the product measure $\mu \otimes \mu$. That is, for any Borel function $h : S \mapsto [0, \infty)$,

$$\mathrm{E}\Big\{ \sum_{(y_i,u_i)\in X} h(y_i, u_i) \Big\} = \int h(y, u)\rho(y, u)\, \mathrm{d}\mu(y, u), \tag{A.1}$$

provided this integral is finite. We say that X is *(first order) homogeneous* if ρ is a constant function. Furthermore, for any Borel function $k : S \times S \mapsto [0, \infty)$,

$$\begin{aligned} &\mathrm{E}\Big\{ \sum_{(y_i,u_i),(y_j,u_j)\in X}^{\neq} k(y_i, u_i, y_j, u_j) \Big\} \\ &= \iint k(y_1, u_1, y_2, u_2)\rho(y_1, u_1)\rho(y_2, u_2)g(y_1, u_1, y_2, u_2)\, \mathrm{d}\mu(y_1, u_1)\, \mathrm{d}\mu(y_2, u_2), \end{aligned} \tag{A.2}$$

provided this double integral is finite. Here, we set $g(y_1, u_1, y_2, u_2) = 0$ if $\rho(y_1, u_1)\rho(y_2, u_2) = 0$, and $\sum^{\neq}_{(y_i,u_i),(y_j,u_j)\in X}$ means that we sum over pairs of distinct points $(y_i, u_i), (y_j, u_j) \in X$.

The functions ρ and g are unique except for null sets with respect to μ and $\mu \otimes \mu$, respectively. For ease of presentation, we ignore null sets in the following. Note that $g(y_1, u_1, y_2, u_2) = g(y_2, u_2, y_1, u_1)$ is symmetric on $S \times S$. We say that X is *stationary in space* if its distribution is invariant under translations of its spatial components; this implies that $\rho(y, u)$ depends only on u, and $g(y_1, u_1, y_2, u_2)$ depends only on (y_1, y_2) through the difference $y_1 - y_2$. If the distribution of X is invariant under rotations (about the origin in $\mathbb{R}^d$) of its spatial components, we say that X is *isotropic in space*. Stationarity and isotropy in space imply that $g(y_1, u_1, y_2, u_2)$ depends only on (y_1, y_2) through the distance $\|y_1 - y_2\|_d$. We say that X is *isotropic on the sphere* if its distribution is invariant under rotations (on $\mathbb{S}^k$) of its spherical components; this implies that $g(y_1, u_1, y_2, u_2)$ depends only on (u_1, u_2) through the geodesic

(great circle/shortest path) distance $d(u_1, u_2)$ on $\mathbb{S}^k$. If X is stationary in space and isotropic on the sphere, then ρ is constant and

$$g(y_1, u_1, y_2, u_2) = g_0\{y_1 - y_2, d(u_1, u_2)\}, \qquad y_1, y_2 \in \mathbb{R}^d,\ u_1, u_2 \in \mathbb{S}^k, \quad \text{(A.3)}$$

depends only on (y_1, y_2) through $y_1 - y_2$ and on (u_1, u_2) through $d(u_1, u_2)$ (this property is studied further in Section 3). If it is furthermore assumed that X is isotropic in space, then

$$g(y_1, u_1, y_2, u_2) = g_*\{\|y_1 - y_2\|_d, d(u_1, u_2)\}, \qquad y_1, y_2 \in \mathbb{R}^d,\ u_1, u_2 \in \mathbb{S}^k,$$

depends only on (y_1, y_2) through $\|y_1 - y_2\|_d$ and on (u_1, u_2) through $d(u_1, u_2)$.

The spatial components of X constitute a usual *spatial point process* $Y = \{y : (y, u) \in X\}$, which is locally finite, whereas the spherical components constitute a *point process on the sphere* $U = \{u : (y, u) \in X\}$ that may be infinite on the compact set $\mathbb{S}^k$. Let $W \subset \mathbb{R}^d$ be a bounded Borel set, which we may think of as a window where the spatial components $Y_W = Y \cap W$ are observed. As X is locally finite, the spherical components associated with Y_W constitute a finite point process $U_W = \{u : (y, u) \in X, y \in W\}$ on $\mathbb{S}^k$. Let $N = N(W)$ denote the cardinality of Y_W. To avoid trivial and undesirable cases, we assume that $|W| > 0$ and that the following inequalities hold:

$$0 < \mathrm{E}(N) < \infty \quad \text{(A.4)}$$

and

$$0 < \mathrm{E}\{N(N-1)\} < \infty, \quad \text{(A.5)}$$

where, by (A.1)–(A.2),

$$\mathrm{E}(N) = \int_{W \times \mathbb{S}^k} \rho(y, u)\, \mathrm{d}\mu(y, u)$$

and

$$\mathrm{E}\{N(N-1)\} = \int_{W \times \mathbb{S}^k} \int_{W \times \mathbb{S}^k} \rho(y_1, u_1)\rho(y_2, u_2) g(y_1, u_1, y_2, u_2)\, \mathrm{d}\mu(y_1, u_1)\, \mathrm{d}\mu(y_2, u_2).$$

Note that Y has intensity function ρ_1 and pair correlation function g_1 given by

$$\rho_1(y) = \int \rho(y, u)\, \mathrm{d}\nu(u), \qquad y \in \mathbb{R}^d, \quad \text{(A.6)}$$

and

$$\rho_1(y_1)\rho_1(y_2)g_1(y_1, y_2) = \iint \rho(y_1, u_1)\rho(y_2, u_2) g(y_1, u_1, y_2, u_2)\, \mathrm{d}\nu(u_1)\, \mathrm{d}\nu(u_2) \quad \text{(A.7)}$$

for $y_1, y_2 \in \mathbb{R}^d$, where we set $g_1(y_1, y_2) = 0$ if $\rho_1(y_1)\rho_1(y_2) = 0$. This follows from (A.1)–(A.2) and definitions of the intensity and pair correlation function for spatial point processes (see e.g. Møller and Waagepetersen, 2004) Clearly, if X is stationary in space, then Y is stationary, ρ_1 is constant, and $g_1(y_1, y_2)$ is stationary, that is, it depends only on $y_1 - y_2$. If in addition X is isotropic in space, then $g_1(y_1, y_2)$ is isotropic, that is, it depends only on $\|y_1 - y_2\|_d$. On the other hand if Y is stationary (or isotropic) and the spherical components are independent of Y, then X is stationary (or isotropic) in space.

Similarly, using definitions of the intensity and pair correlation function for point processes on the sphere (Lawrence et al., 2016; Møller and Rubak, 2016), U_W has intensity function ρ_2 (with respect to ν) and pair correlation function g_2 (with respect to $\nu \otimes \nu$) given by

$$\rho_2(u) = \int_W \rho(y, u) \, \mathrm{d}y, \qquad u \in \mathbb{S}^k, \tag{A.8}$$

and

$$\begin{aligned} &\rho_2(u_1)\rho_2(u_2)g_2(u_1, u_2) \\ &\qquad = \int_W \int_W \rho(y_1, u_1)\rho(y_2, u_2)g(y_1, u_1, y_2, u_2) \, \mathrm{d}y_1 \, \mathrm{d}y_2 \end{aligned} \tag{A.9}$$

for $u_1, u_2 \in \mathbb{S}^k$, where we set $g_2(u_1, u_2) = 0$ if $\rho_2(u_1)\rho_2(u_2) = 0$. Note that we suppress in the notation that ρ_2 and g_2 depend on W. Obviously, if X is isotropic on the sphere, then ρ_2 is constant and $g_2(u_1, u_2)$ is isotropic as it depends only on $d(u_1, u_2)$.

2.2 Examples

The following examples introduce the point process models considered in this paper.

Example 1 (Poisson and Cox processes). First, suppose X is a *Poisson process* with a locally integrable intensity function ρ. This means that, the count $N(A) = \#X_A$ is Poisson distributed with mean $\int_A \rho(y, u) \, \mathrm{d}\mu(y, u)$ for any bounded Borel set $A \subset S$ and, conditional on $N(A)$, the points in X_A are independent and identically distributed (IID) with a density proportional to ρ restricted to A. Note that $g = 1$. Further, X is stationary in space and isotropic on the sphere if and only if ρ is constant, in which case we call X a *homogeneous Poisson process* with intensity ρ. Furthermore, Y and U_W are Poisson processes, so $g_1 = 1$ and $g_2 = 1$.

Second, let $\Lambda = \{\Lambda(y, u) : (y, u) \in S\}$ be a non-negative random field so that with probability one $\int_A \Lambda(y, u) \, \mathrm{d}\mu(y, u)$ is finite for any bounded Borel set $A \subset S$. If X conditioned on Λ is a Poisson process with intensity function

Λ, then X is said to be a *Cox process driven by* Λ (Cox, 1955). Clearly, the intensity and pair correlation functions of X are

$$\rho(y,u) = \mathrm{E}\{\Lambda(y,u)\}, \qquad y \in \mathbb{R}^d,\ u \in \mathbb{S}^k, \tag{A.10}$$

and

$$\rho(y_1,u_1)\rho(y_2,u_2)g(y_1,u_1,y_2,u_2) = \mathrm{E}\{\Lambda(y_1,u_1)\Lambda(y_2,u_2)\},$$

for $y_1, y_2 \in \mathbb{R}^d$ and $u_1, u_2 \in \mathbb{S}^k$. To separate the intensity function ρ from random effects, it is convenient to work with a so-called residual random field $R = \{R(y,u) : (y,u) \in S\}$ fulfilling $\Lambda(y,u) = \rho(y,u)R(y,u)$, so $\mathrm{E}\{R(y,u)\} = 1$ (see e.g. Møller and Waagepetersen, 2007; Diggle, 2014). Then

$$g(y_1,u_1,y_2,u_2) = \mathrm{E}\{R(y_1,u_1)R(y_2,u_2)\}, \qquad y_1,y_2 \in \mathbb{R}^d,\ u_1,u_2 \in \mathbb{S}^k, \tag{A.11}$$

whenever $\rho(y_1,u_1)\rho(y_2,u_2) > 0$.

Note that projected point processes Y and U_W are Cox processes driven by the random fields $\{\int_{\mathbb{S}^k} \Lambda(y,u)\,\mathrm{d}\nu(u) : y \in \mathbb{R}^d\}$ and $\{\int_W \Lambda(y,u)\,\mathrm{d}y : u \in \mathbb{S}^k\}$, respectively. Their intensity and pair correlation functions are specified by (A.6)–(A.9).

Example 2 (Log Gaussian Cox processes). A Cox process X is called a *log Gaussian Cox process* (*LGCP*; Møller et al., 1998) if the residual random field is of the form $R = \exp(Z)$, where Z is a Gaussian random field (GRF) with mean function $\mu(y,u) = -c(y,u,y,u)/2$, where c is the covariance function of Z. Note that X has pair correlation function

$$g(y_1,u_1,y_2,u_2) = \exp\left[c\{(y_1,u_1),(y_2,u_2)\}\right] \tag{A.12}$$

for $y_1, y_2 \in \mathbb{R}^d$ and $u_1, u_2 \in \mathbb{S}^k$.

Example 3 (Marked point processes). It is sometimes useful to view X as a marked point process (see e.g. Daley and Vere-Jones, 2003; Illian et al., 2008), where the spatial components are treated as the ground process and the spherical components as marks. Often it is of interest to test the hypothesis H_0 that the marks are IID and independent of the ground process Y. Under H_0, with each mark following a density p with respect to ν, the intensity is

$$\rho(y,u) = \rho_1(y)p(u), \qquad y \in \mathbb{R}^d,\ u \in \mathbb{S}^k,$$

and the pair correlation function

$$g(y_1,u_1,y_2,u_2) = g_1(y_1,y_2), \qquad y_1,y_2 \in \mathbb{R}^d,\ u_1,u_2 \in \mathbb{S}^k,$$

does not depend on (u_1,u_2).

In some situations, it may be more natural to look at it conversely, that is, treating U_W as the ground process and Y_W as marks. Then similar results for ρ and g may be established by interchanging the roles of points and marks.

Example 4 Independently marked determinantal point processes. Considering a space-sphere point process X as a marked point process that fulfils the hypothesis H_0 given in Example 3, we may let the ground process Y be distributed according to any point process model of our choice regardless of the marks U. For instance, in case of repulsion between the points in Y, a *determinantal point process (DPP)* may be of interest because of its attractive properties (see Lavancier et al., 2015, and the references therein). Briefly, a DPP is defined by a so-called kernel $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, which we assume is a complex covariance function, that is, C is positive semi-definite and Hermitian. Furthermore, let $\rho_1^{(n)}$ denote the nth order joint intensity function of Y, that is, $\rho_1^{(1)} = \rho_1$ is the intensity and $\rho_1^{(2)}(y_1, y_2) = \rho_1(y_1)\rho_1(y_2)g_1(y_1, y_2)$ for $y_1, y_2 \in \mathbb{R}^d$, while we refer to Lavancier et al. (2015) for the general definition of $\rho_1^{(n)}$ which is an extension of (A.6)–(A.7). If for all $n = 1, 2, \ldots,$

$$\rho_1^{(n)}(y_1, \ldots, y_n) = \det\{C(y_i, y_j)\}_{i,j=1,\ldots,n}, \qquad y_1, \ldots, y_n \in \mathbb{R}^d,$$

where $\det\{C(y_i, y_j)\}_{i,j=1,\ldots,n}$ is the determinant of the $n \times n$ matrix with (i, j)-entry $C(y_i, y_j)$, we call Y a DPP with kernel C and refer to X as an *independently marked DPP*. It follows that Y has intensity function $\rho(y) = C(y, y)$ and pair correlation function

$$g_1(y_1, y_2) = 1 - |R(y_1, y_2)|^2, \qquad y_1, y_2 \in \mathbb{R}^d,$$

whenever $\rho(y_1)\rho(y_2) > 0$, where $R(y_1, y_2) = C(y_1, y_2)/\sqrt{C(y_1, y_1)C(y_2, y_2)}$ is the correlation function corresponding to C and $|z|$ denotes the modulus of $z \in \mathbb{C}$.

Alternatively, we may look at a DPP on the sphere (Møller et al., 2018), that is, modelling U_W as a DPP while considering Y_W as the marks and impose the conditions of IID marks independent of U_W.

3 The space-sphere K-function

3.1 Definition

When (A.3) holds we say that the space-sphere point process X is *second order intensity-reweighted stationary* (SOIRS) and define the *space-sphere K-function* by

$$K(r, s) = \int_{\|y\|_d \le r,\, d(u,e) \le s} g_0\{y, d(u, e)\}\, \mathrm{d}\mu(y, u), \qquad r \ge 0,\ 0 \le s \le \pi, \quad \text{(A.13)}$$

where $e \in \mathbb{S}^k$ is an arbitrary reference direction. This definition does not depend on the choice of e, as the integrand only depends on $u \in \mathbb{S}^k$ through its

geodesic distance to e and $\nu(\cdot)$ is a rotation invariant measure. For example, we may let $e = (0, \ldots, 0, 1) \in \mathbb{S}^k$ be the "North Pole".

Let $\sigma_k = \nu(\mathbb{S}^k) = 2\pi^{(k+1)/2}/\Gamma\{(k+1)/2\}$ denote the surface measure of $\mathbb{S}^k$. For any Borel set $B \subset \mathbb{R}^d$ with $0 < |B| < \infty$, we easily obtain from (A.2) and (A.13) that

$$
\begin{aligned}
&K(r,s) \\
&= \frac{1}{|B|\sigma_k} \iint_{y_1 \in B,\, \|y_1 - y_2\|_d \le r,\, d(u_1,u_2) \le s} g_0\{y_1 - y_2, d(u_1,u_2)\}\, \mathrm{d}\mu(y_1,u_1)\, \mathrm{d}\mu(y_2,u_2) \\
&= \frac{1}{|B|\sigma_k} \mathrm{E}\Big[\sum_{(y_i,u_i),(y_j,u_j) \in X}^{\neq} \frac{\mathbb{I}\{y_i \in B,\, \|y_i - y_j\|_d \le r,\, d(u_i,u_j) \le s\}}{\rho(y_i,u_i)\rho(y_j,u_j)} \Big] \qquad \text{(A.14)}
\end{aligned}
$$

for $r \ge 0$, $0 \le s \le \pi$, where $\mathbb{I}(\cdot)$ denotes the indicator function. The relation given by (A.14) along with the requirement that the expression in (A.14) does not depend on the choice of B could alternatively have been used as a more general definition of the space-sphere K-function. Such a definition is in agreement with the one used in Baddeley et al. (2000) for SOIRS of a spatial point process. It is straightforward to show that (A.14) does not depend on B when X is stationary in space.

For $r, s > 0$ and $(y_1, u_1), (y_2, u_2) \in S$, we say that (y_1, u_1) and (y_2, u_2) are (r,s)-close neighbours if $\|y_1 - y_2\|_d \le r$ and $d(u_1, u_2) \le s$. If X is stationary in space and isotropic on the sphere, then (A.14) shows that $\rho K(r,s)$ can be interpreted as the expected number of further (r,s)-close neighbours in X of a typical point in X. More formally, this interpretation relates to the reduced Palm distribution (Daley and Vere-Jones, 2003).

Some literature treating marked point processes discuss the so-called *mark-weighted K-function* (see e.g. Illian et al., 2008; Koubek et al., 2016), which to some extent resembles the space-sphere K-function in a marked point process context; both are cumulative second order summary functions that consider points as well as marks. However, the mark-weighted K-function has an emphasis on the marked point process setup (and considers e.g. ρ_1 rather than ρ), whereas the space-sphere K-function is constructed in such a way that it is an analogue to the planar/spherical K-function for space-sphere point processes.

Example 1 continued (Poisson and Cox processes). A Poisson process is clearly SOIRS and $K(r,s)$ is simply the product of the volume of a d-dimensional ball with radius r and the surface area of a spherical cap given by $\{u \in \mathbb{S}^k : d(u,e) \le s\}$ for an arbitrary $e \in \mathbb{S}^k$ (see Li, 2011, for formulas of this area). Thus, for $r \ge 0$, the space-sphere K-function is

$$
K_{Pois}(r,s) = \begin{cases} \frac{r^d \pi^{(d+k+1)/2}}{\Gamma(1+d/2)\Gamma\{(k+1)/2\}} I_{\sin^2(s)}\left(\frac{k}{2}, \frac{1}{2}\right), & 0 \le s \le \frac{\pi}{2}, \\ \frac{r^d \pi^{(d+k+1)/2}}{\Gamma(1+d/2)\Gamma\{(k+1)/2\}} \{2 - I_{\sin^2(\pi-s)}(\frac{k}{2}, \frac{1}{2})\}, & \frac{\pi}{2} < s \le \pi, \end{cases}
$$

where $I_x(a,b)$ is the regularized incomplete beta function. In particular, if $k=2$,

$$\left.\begin{array}{ll} I_{\sin^2(s)}(\frac{k}{2},\frac{1}{2}), & 0 \le s \le \frac{\pi}{2} \\ 2 - I_{\sin^2(\pi-s)}(\frac{k}{2},\frac{1}{2}), & \frac{\pi}{2} < s \le \pi \end{array}\right\} = 1-\cos(s).$$

If the residual random field R in (A.11) is invariant under translations in $\mathbb{R}^d$ and under rotations on $\mathbb{S}^k$, then the associated Cox process is SOIRS. The evaluation of g (and thus K) depends on the particular model of R as exemplified in Example 2 below and in Section 7.

Example 2 continued (LGCPs). Suppose that the distribution of R is invariant under translations in $\mathbb{R}^d$ and under rotations on $\mathbb{S}^k$, and recall that R is required to have unit mean. Then the underlying GRF Z has a covariance function of the form

$$c(y_1,u_1,y_2,u_2) = c_0\{y_1-y_2, d(u_1,u_2)\}, \qquad y_1,y_2 \in \mathbb{R}^d,\ u_1,u_2 \in \mathbb{S}^k,$$

and $\mathrm{E}Z(y,u) = -\sigma^2/2$ for all $y \in \mathbb{R}^d$ and $u \in \mathbb{S}^k$, where $\sigma^2 = c_0(0,0)$ is the variance. It then follows from (A.12) that X is SOIRS with

$$g_0(y,s) = \exp\{c_0(y,s)\}, \qquad y \in \mathbb{R}^d,\ 0 \le s \le \pi. \tag{A.15}$$

4 Separability

4.1 First order separability

We call the space-sphere point process X *first order separable* if there exist non-negative Borel functions f_1 and f_2 such that

$$\rho(y,u) = f_1(y)f_2(u), \qquad y \in \mathbb{R}^d,\ u \in \mathbb{S}^k.$$

By (A.4), (A.6), and (A.8) this is equivalent to

$$\rho(y,u) = \rho_1(y)\rho_2(u)/\mathrm{E}(N), \qquad y \in \mathbb{R}^d,\ u \in \mathbb{S}^k, \tag{A.16}$$

recalling that ρ_2 and N depend on W, but $\rho_2/\mathrm{E}N$ does not depend on the choice of W. Then, in a marked point process setup where the spherical components are treated as marks, $\rho_2(\cdot)/\mathrm{E}(N)$ is the density of the mark distribution. First order separability was seen in Example 3 to be fulfilled under the assumption of IID marks independent of the ground process. Moreover, any homogeneous space-sphere point process is clearly first order separable. In practice, first order separability is a working hypothesis which may be hard to check.

4.2 Second order separability

If there exist Borel functions k_1 and k_2 such that

$$g(y_1, u_1, y_2, u_2) = k_1(y_1, y_2)k_2(u_1, u_2), \qquad y_1, y_2 \in \mathbb{R}^d,\ u_1, u_2 \in \mathbb{S}^k,$$

we call X *second order separable*. Assuming first order separability, it follows by (A.5), (A.7), (A.9), and (A.16) that second order separability is equivalent to

$$g(y_1, u_1, y_2, u_2) = \beta g_1(y_1, y_2)g_2(u_1, u_2), \qquad y_1, y_2 \in \mathbb{R}^d,\ u_1, u_2 \in \mathbb{S}^k, \quad \text{(A.17)}$$

where

$$\beta = \mathrm{E}(N)^2 / \mathrm{E}\{N(N-1)\}$$

and noting that β and g_2 depend on W, but βg_2 does not depend on the choice of W. The value of β may be of interest: for a Poisson Process, $\beta = 1$; for a Cox process, $\mathrm{var}(N) \geq \mathrm{E}(N)$ (see e.g. Møller and Waagepetersen, 2004), so $\beta \leq 1$; for an independently marked DPP, $\beta \geq 1$ (Lavancier et al., 2015).

Example 1 continued (Poisson and Cox processes). Clearly, when X is a Poisson process, it is second order separable. Assume instead that X is a Cox process and the residual random field is separable, that is, $R(y, u) = R_1(y)R_2(u)$, where $R_1 = \{R_1(y) : y \in \mathbb{R}^d\}$ and $R_2 = \{R_2(u) : u \in \mathbb{S}^k\}$ are independent random fields. Then, by (A.11), X is second order separable and

$$g(y_1, u_1, y_2, u_2) = \mathrm{E}\{R_1(y_1)R_1(y_2)\}\mathrm{E}\{R_2(u_1)R_2(u_2)\}$$

for $y_1, y_2 \in \mathbb{R}^d$ and $u_1, u_2 \in \mathbb{S}^k$.

Example 2 continued (LGCPs). If X is a LGCP driven by $\Lambda(y, u) = \rho(y, u) \cdot \exp\{Z(y, u)\}$, second order separability is implied if $Z_1 = \log R_1$ and $Z_2 = \log R_2$ are independent GRFs so that $Z(y, u) = Z_1(y) + Z_2(u)$. Then, by the imposed invariance properties of the distribution of the residual random field, Z_1 must be stationary with a stationary covariance function $c_1(y_1, y_2) = c_{01}(y_1 - y_2)$ and mean $-c_{01}(0)/2$, and Z_2 must be isotropic with an isotropic covariance function $c_2(u_1, u_2) = c_{02}\{d(u_1, u_2)\}$ and mean $-c_{02}(0)/2$. Consequently, in (A.15), $c_0(y, s) = c_{01}(y) + c_{02}(s)$ for $y \in \mathbb{R}^d$ and $0 \leq s \leq \pi$.

Example 3 continued (marked point processes). Consider the space-sphere point process X as a marked point process with marks in $\mathbb{S}^k$. As previously seen, first and second order separability is fulfilled under the assumption of IID marks independent of the ground process, but we may in fact work with weaker conditions to ensure the separability properties as follows. Assume that each mark is independent of the ground process Y and the marks are identically distributed following a density function p with respect to ν. Then

the first order separability condition (A.16) is satisfied with $\rho_2(u) = \mathrm{E}(N)p(u)$ for $u \in \mathbb{S}^k$. In addition, assuming the conditional distribution of the marks given Y is such that any pair of marks is independent of Y and follows the same joint density $q(\cdot,\cdot)$ with respect to $\nu \otimes \nu$, it is easily seen that the second order separability condition (A.17) is satisfied with

$$g_2(u_1,u_2) = \frac{q(u_1,u_2)}{\beta p(u_1)p(u_2)}, \qquad u_1,u_2 \in \mathbb{S}^k,$$

whenever $\rho_2(u_1)\rho_2(u_2) > 0$. If we also have pairwise independence between the marks, that is, $q(u_1,u_2) = p(u_1)p(u_2)$, then the pair correlation function $g(y_1,u_1,y_2,u_2) = g_1(y_1,y_2)$ does not depend on (u_1,u_2) and $g_2(u_1,u_2) = 1/\beta$ is constant. Note that this implies $g_2 \leq 1$ for an independently marked DPP, reflecting that even when the marks are drawn independently of Y the behaviour of the points implicitly affects the marks as the number of points is equal to the number of marks.

Again, the roles of points and marks may be switched resulting in statements analogue to those above.

4.3 Assuming both SOIRS and first and second order separability

Suppose that X is both SOIRS and first and second order separable. Then the space-sphere K-function can be factorized as follows. Note that Y and U_W are SOIRS since there by (A.3), (A.7), (A.9), and (A.16) exist Borel functions g_{01} and g_{02} such that

$$\begin{aligned} g_1(y_1,y_2) &= g_{01}(y_1-y_2) \\ &= \iint \frac{\rho_2(u_1)}{\mathrm{E}(N)}\frac{\rho_2(u_2)}{\mathrm{E}(N)} g_0\{y_1-y_2, d(u_1,u_2)\}\,\mathrm{d}\nu(u_1)\,\mathrm{d}\nu(u_2) \end{aligned} \tag{A.18}$$

for $y_1,y_2 \in \mathbb{R}^d$ with $\rho_1(y_1)\rho_1(y_2) > 0$, and

$$\begin{aligned} g_2(u_1,u_2) &= g_{02}\{d(u_1,u_2)\} \\ &= \int_W \int_W \frac{\rho_1(y_1)}{\mathrm{E}(N)}\frac{\rho_1(y_2)}{\mathrm{E}(N)} g_0\{y_1-y_2, d(u_1,u_2)\}\,\mathrm{d}y_1\,\mathrm{d}y_2 \end{aligned} \tag{A.19}$$

for $u_1,u_2 \in \mathbb{S}^k$ with $\rho_2(u_1)\rho_2(u_2) > 0$. Hence, the inhomogeneous K-function for the spatial components in Y (introduced in Baddeley et al., 2000) is

$$K_1(r) = \int_{\|y\|_d \leq r} g_{01}(y)\,\mathrm{d}y, \qquad r \geq 0,$$

and the inhomogeneous K-function for the spherical components in U_W (introduced in Lawrence et al., 2016; Møller and Rubak, 2016) is

$$K_2(s) = \int_{d(u,e)\leq s} g_{02}\{d(u,e)\}\,\mathrm{d}\nu(u), \qquad 0 \leq s \leq \pi,$$

where $e \in \mathbb{S}^k$ is arbitrary. Combining (A.13) and (A.17)–(A.19), we obtain

$$K(r,s) = \beta K_1(r) K_2(s), \qquad r \geq 0,\, 0 \leq s \leq \pi.$$

Note that, if X is a first order separable Poisson process, then $D(r,s) = K(r,s) - K_1(r)K_2(s)$ is 0, and an estimate of D may also be used as a functional summary statistic when testing a Poisson hypothesis.

5 Estimation of K-functions

In this section, we assume for specificity that the observation window is $W \times \mathbb{S}^k$, where $W \subset \mathbb{R}^d$ is a bounded Borel set, and a realisation $X_{W\times\mathbb{S}^k} = x_{W\times\mathbb{S}^k}$ is observed; in Section 7, we discuss other cases of observation windows. We let $Y_W = y_W$ and $U_W = u_W$ be the corresponding sets of observed spatial and spherical components.

First, assume that ρ_1 and ρ_2 are known. Following Baddeley et al. (2000), we estimate K_1 by

$$\hat{K}_1(r) = \sum_{y_i,y_j\in y_W}^{\neq} \frac{\mathbb{I}(\|y_i - y_j\|_d \leq r)}{w_1(y_i,y_j)\rho_1(y_i)\rho_1(y_j)}, \qquad r \geq 0, \tag{A.20}$$

where w_1 is an edge correction factor on $\mathbb{R}^d$. If we let $w_1(y_i,y_j) = |W \cap W_{y_i-y_j}|$ be the translation correction factor (Ohser, 1983), where $W_y = \{y+z : z \in W\}$ denotes the translation of W by $y \in \mathbb{R}^d$, then $\hat{K}_1$ is an unbiased estimate of K_1 (see e.g. Lemma 4.2 in Møller and Waagepetersen, 2004). For $d = 1$, we may instead use the temporal edge correction factor with $w_1(y_i,y_j) = |W|$ if $[y_i - y_j, y_i + y_j] \subseteq W$ and $w_1(y_i,y_j) = |W|/2$ otherwise (Diggle et al., 1995; Møller and Ghorbani, 2012). Moreover, for estimation of K_2, we use the unbiased estimate

$$\hat{K}_2(s) = \frac{1}{\sigma_k} \sum_{u_i,u_j\in u_W}^{\neq} \frac{\mathbb{I}\{d(u_i,u_j) \leq s\}}{\rho_2(u_i)\rho_2(u_j)}, \qquad 0 \leq s \leq \pi, \tag{A.21}$$

cf. Lawrence et al. (2016) and Møller and Rubak (2016). A natural extension of the above estimates gives the following estimate of K:

$$\hat{K}(r,s) = \frac{1}{\sigma_k} \sum_{(y_i,u_i),(y_j,u_j)\in x_{W\times\mathbb{S}^k}}^{\neq} \frac{\mathbb{I}\{\|y_i - y_j\|_d \leq r,\, d(u_i,u_j) \leq s\}}{w_1(y_i,y_j)\rho(y_i,u_i)\rho(y_j,u_j)} \tag{A.22}$$

for $r \geq 0$, $0 \leq s \leq \pi$. This is straightforwardly seen to be an unbiased estimate when w_1 is the translation correction factor.

Second, in practice we need to replace ρ_1 in (A.20), ρ_2 in (A.21), and ρ in (A.22) by estimates, as exemplified in Section 6. This may introduce a bias.

6 Data examples

6.1 Fireball locations over time

Figure A.2 shows the time and location of $n = 344$ fireballs observed over a time period from `2005-01-01 03:44:09` to `2016-08-12 23:59:59` corresponding to a time frame W of about 606 weeks. The data can be recovered at `http://neo.jpl.nasa.gov/fireballs/` using these time stamps. Figure A.2 reveals no inhomogeneity of neither fireball locations or event times. Therefore we assumed first order homogeneity, and used the following unbiased estimates for the intensities:

$$\hat{\rho}_1 = n/|W| = 0.57, \;\; \hat{\rho}_2 = n/(4\pi) = 27.37, \;\; \hat{\rho} = n/(4\pi|W|) = 0.05.$$

Then $\hat{K}_1$, $\hat{K}_2$, and $\hat{K}$ (with w_1 in (A.20) and (A.22) equal to the temporal edge correction factor) were used as test functions in three different global rank envelope tests for testing whether fireball event times, locations, and locations over time each could be described by a homogeneous Poisson model with estimated intensity $\hat{\rho}_1$, $\hat{\rho}_2$, and $\hat{\rho}$, respectively. Appendix A provides a brief account on global rank envelope tests; see also Myllymäki et al. (2017). Under each of the three fitted Poisson processes and using 2499 simulations (as recommended in Myllymäki et al., 2017), we obtained p-intervals of $(0.028, 0.040)$ for the event times, $(0.908, 0.908)$ for the locations, and $(0.445, 0.516)$ for the locations over time. The associated 95% global rank envelopes for $\hat{K}_1$ and $\hat{K}_2$ are shown in Figure A.3, and the difference between $\hat{K}$ and the upper and lower 95% global rank envelope is shown in Figure A.4. Since $\hat{K}_2$ and $\hat{K}$ stay inside the 95% global rank envelopes for the considered distances on $\mathbb{S}^k$ and $\mathbb{R} \times \mathbb{S}^k$, there is no evidence against a homogeneous Poisson model for neither locations or locations over time. On the other hand, with a conservative p-value of 4%, the global rank envelope test based on $\hat{K}_1$ indicates that a homogeneous Poisson model for the event times is not appropriate. However, the observed test function $\hat{K}_1(r)$ falls only outside the envelope in Figure A.3 for large values of r. Thus, choosing a slightly smaller interval of r-values would lead to a different conclusion.

As an alternative to the space-sphere K-function, we considered the summary function $D(r, s)$ which in case of a Poisson process is 0. Estimating D by $\hat{D}(r, s) = \hat{K}(r, s) - \hat{K}_1(r)\hat{K}_2(s)$, we performed a global rank envelope test with D as test function. The resulting test gave a p-interval of $(0.537, 0.564)$ which is similar to the one obtained using $\hat{K}$ as test function.

6.2 Location and orientation of pyramidal neurons

We now return to the space-sphere point pattern concerning location and orientation of pyramidal neurons described in Section 1, which is a data set

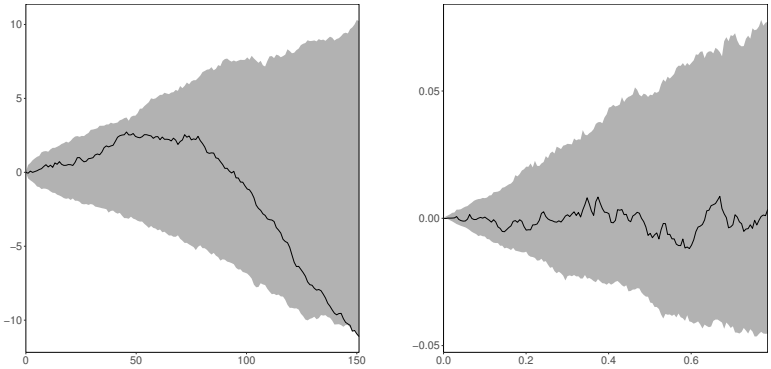

Fig. A.3: Left: $\hat{K}_1(r) - 2r$ for the fireball event times (solid curve) along with a 95% global rank envelope (grey area) under a homogeneous Poisson model on the time interval for the observed events. Right: $\hat{K}_2(s) - 2\pi\{1 - \cos(s)\}$ for the fireball locations (solid curve) along with a 95% global rank envelope (grey area) under a homogeneous Poisson model on $\mathbb{S}^2$.

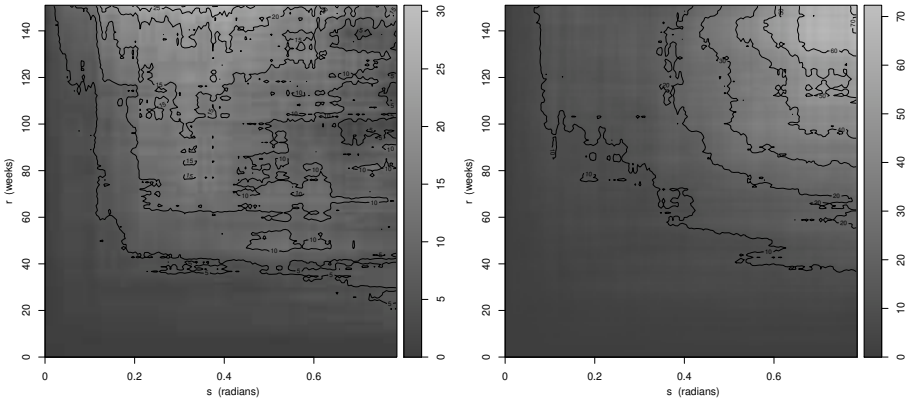

Fig. A.4: Difference between $\hat{K}(r,s)$ for the observed fireball locations over time and the lower ($\hat{K}_{\text{low}}$) or upper ($\hat{K}_{\text{upp}}$) 95% global rank envelope under a homogeneous Poisson model on $W \times \mathbb{S}^2$. Left: $\hat{K}(r,s) - \hat{K}_{\text{low}}(r,s)$. Right: $\hat{K}_{\text{upp}}(r,s) - \hat{K}(r,s)$.

collected by Ali H. Rafati, a biomedical and clinical scientist. The point pattern is observed on $W \times \mathbb{S}^2$, where $W \subset \mathbb{R}^3$ is the rectangular box shown in Figure A.1 with side lengths 492.7 µm, 132.0 µm, and 407.7 µm. Due to the way data was collected, 46 neurons in the original dataset had an orientation/unit vector lying exactly in the x-z plane, meaning that the orientations of these 46 neurons are located only on a great circle of $\mathbb{S}^2$. To keep the analysis simple we disregarded these neurons, resulting in a dataset consisting of $n = 504$ neurons. Below we initially discuss an appropriate parametric forms of the intensity for the locations and orientations and how to estimate the intensity parameters. Then we investigate whether the orientations and locations can be described by a Poisson model with the proposed intensity, where we first consider the data as two separate point patterns (a spatial point pattern describing the locations and a spherical point pattern describing the orientations) and next as a space-sphere point pattern.

Figure A.1 reveals no inhomogeneity for the neuron locations, whereas it is evident that the orientations are inhomogeneous pointing mostly toward the pial surface of the brain (the plane perpendicular to the z-axis). Thus, we estimated the intensity of the locations by $\hat{\rho}_1 = n/|W| = 1.9 \times 10^{-5}$, and for the orientations we let the intensity be $\rho_2(u) = nf(u)$, where f is a density on $\mathbb{S}^2$ which we model as follows. Figure A.5 indicates that the orientations arise from a mixture of two distributions; one distribution with points falling close to the North Pole and another with points falling in a narrow girdle. Therefore, we let $f(u) = pf_K(u) + (1-p)f_W(u)$ be the mixture density of a Kent and a Watson distribution on $\mathbb{S}^2$ (see e.g. Fisher et al., 1987) for a detailed description of these spherical distributions). In brief, the Kent density, f_K, depends on five parameters (three directional, one concentration, and one ovalness parameter), and its contours are oval with centre and form specified by the directional parameters. Depending on the values of the ovalness and concentration parameter, the Kent distribution is either uni- or bimodal. Here, to account for the large number of points centred around the North Pole, we consider the unimodal Kent distribution. Furthermore, the Watson density, f_W, depends on two parameters; a directional parameter determining the centres of the density's circular contours, and a concentration parameter controlling where and how fast the density peaks. Depending on the sign of the concentration parameter, the density either decreases or increases as the geodesic distance to the centres of the contours increases, giving rise to either a bimodal or girdle shaped distribution. Since the Watson distribution shall describe the orientations on the girdle, the concentration parameter must be negative.

The eight parameters of the proposed intensity function ρ_2 were estimated as follows. The orientations occurring on the southern hemisphere are presumed to come from the Watson distribution, while the orientations on the northern hemisphere come from both distributions. Therefore, and

because the Watson density on the northern hemisphere is a reflection of the southern hemisphere, we simple estimated the mixture probability by $\hat{p} = 1 - 2n_s/n = 0.94$, where n_s is the observed number of points on the southern hemisphere. The directional parameters were chosen based on expectations expressed by the scientist behind the data collections, which were supported by visual inspection of the data; e.g. the directional parameter for the Kent distribution that determines the centre of the contours was chosen as the North Pole corresponding to the direction perpendicular to the pial surface and consistent with Figure A.5. Finally, the concentration and ovalness parameters were estimated by numerical maximization of the profile likelihood, giving the estimated density

$$\hat{f}(u) = 0.94\, C_K \exp\{14.89u_3 + 2.69(u_1^2 - u_2^2)\} + 0.06\, C_W \exp(-7.88u_2^2),$$

where $u = (u_1, u_2, u_3) \in \mathbb{S}^2$ and C_K, C_W are normalising constants (see Fisher et al., 1987, for details). Figure A.5 suggests that the fitted density (and associated marginal densities found by numerical integration of $\hat{f}$) adequately describe the distribution of the observed orientations. Therefore, we now turn to investigate whether the locations and orientations can be described by Poisson models with the estimated intensities.

First, we considered the locations and orientations separately and used $\hat{K}_1$ and $\hat{K}_2$, respectively, as test functions for the global rank envelope procedure. Using 2499 simulations from a homogeneous Poisson process on W, we obtained a global rank envelope test with p-interval $(0.000, 0.020)$. Thus, we reject that the locations follow a homogeneous Poisson model. The associated 95% global rank envelopes in Figure A.6 show that the rejection is due to the observed $\hat{K}_1(r)$ falling below the envelope for r-values between 10 µm to 25 µm. This suggests that the observed locations exhibit some degree of regularity that needs to be modelled. For the orientations, a global rank envelope test based on 2499 simulations under an inhomogeneous Poisson model on $\mathbb{S}^2$ with intensity $\hat{\rho}_2(u) = n\hat{f}(u)$ gave a p-interval of $(0.475, 0.481)$ and thus no evidence against the proposed model. Figure A.6 shows the associated 95% global rank envelope.

Second, we considered the data as a space-sphere point pattern and used $\hat{K}$ and $\hat{D}$ to test for a Poisson model with a separable intensity estimated by $\hat{\rho}(y, u) = \hat{\rho}_1\hat{\rho}_2(u)/n = n\hat{f}(u)/|W|$, cf. (A.16). As the test functions $\hat{K}(r, s)$ and $\hat{D}(r, s)$ depend on the two-dimensional argument (r, s) and they are non-smooth with large jumps due to few orientations occurring in areas with low intensity, we increased the number of simulations to 49999 in order to improve the quality of the global rank envelope test (49999 is a much higher number than recommended in Myllymäki et al., 2017, for test functions depending on one argument only). This resulted in the p-intervals $(0.547, 0.549)$ for $\hat{K}$ and $(0.000, 0.003)$ for $\hat{D}$; plots of the difference between the associated

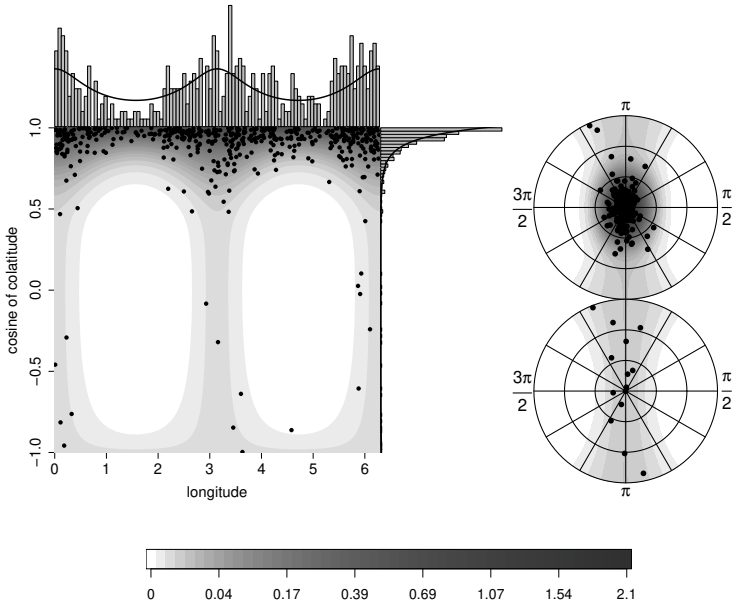

Fig. A.5: Plots of the observed neuron orientations (dots) and the fitted mixture density $\hat{f}$ (grey scale). Left: plot of orientations represented as cosine-colatitude and longitude along with their marginal histograms and marginal fitted densities (solid curves) found by numerically integrating $\hat{f}$. Right: stereographic projection of the northern (top) and southern (bottom) hemisphere.

envelopes and the observed test function are shown in Figure A.7. In conclusion, the test based on $\hat{K}$ reveals no evidence against the proposed space-sphere Poisson model even though the corresponding Poisson model for the locations was rejected by the test based on $\hat{K}_1$. However, the test based on $\hat{D}$ provides a great deal of evidence against the model. This conclusion is probably due to the fact that for this data set $\hat{K}_1(r)\hat{K}_2(s) \gg \hat{K}(r,s)$, meaning that the test based on $\hat{D}$ is highly controlled by the values of $\hat{K}_1$ and $\hat{K}_2$, which results in a rejection for r-values from 10 µm to 20 µm, in line with the test based on $\hat{K}_1$.

It is unsatisfactory that $\hat{K}$ does not detect any deviation from Poisson when $\hat{K}_1$ clearly does, but we expect that the large jumps in $\hat{K}(r,s)$, caused by (r,s)-close neighbours with low intensity, may explain why no evidence against the model is detected. The few orientations that were modelled using a Watson distribution mostly fall in places with very low intensity. Therefore, we independently thinned the space-sphere point pattern with retention probability $\hat{p}\hat{f}_K(u)/\hat{f}(u)$ for $u \in \mathbb{S}^2$. Thereby (with high probability) we removed neurons with orientations that were most likely generated by the Watson distribution. This lead to removal of 26 neurons. For the thinned

data, the global rank envelope test based on $\hat{K}$ for testing the hypothesis of an inhomogeneous Poisson process with intensity proportional to a Kent density gave a p-interval of $(0.052, 0.058)$. Still, the model was not rejected at a 5% significance level, but we at least got closer to a rejection; and so we continued the analysis with the thinned data. The analysis here indicates that, at least in some cases, the power of the global rank envelope test based on $\hat{K}$ may be small. This is investigated further in Section 6.3.

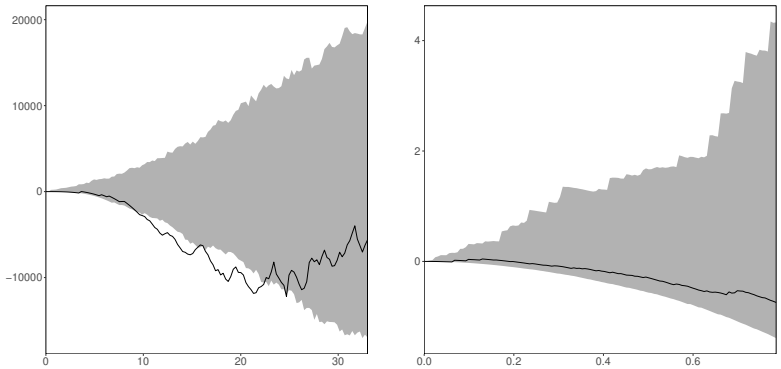

Fig. A.6: Left: $\hat{K}_1(r) - 4\pi r^3/3$ for the neuron locations (solid curve) along with a 95% global rank envelope (grey area) under a homogeneous Poisson model on $W \subseteq \mathbb{R}^3$. Right: $\hat{K}_2(s) - 2\pi\{1 - \cos(s)\}$ for the neuron orientations (solid curve) along with a 95% global rank envelope (grey area) under the fitted inhomogeneous Poisson model on $\mathbb{S}^2$.

As we have seen, a homogeneous Poisson model is not adequate for the locations, and thus a Poisson model with intensity $\hat{\rho}$ as described above is not suitable for describing the space-sphere point pattern. To investigate whether orientations and locations can be modelled separately, that is, whether the locations and orientations are independent, we kept the locations fixed, and independent of the locations we simulated IID orientations from the fitted Kent distribution. The resulting global rank envelope test based on 49999 of such simulations gave a p-interval of $(0.9255, 0.9258)$ for $\hat{K}$ and $(0.1265, 0.1266)$ for $\hat{D}$, showing no evidence against the hypothesis of independence between locations and orientations. Alternatively, if one does not have a suitable model to simulate the spherical (or spatial) components from, the independence test may be performed by randomly permuting the components. Formally, this tests only the hypothesis of exchangeability; a property that is fulfilled under independence. Performing such a permutation test for our data where the locations were fixed and the orientations permuted 49999 times resulted in a p-interval of $(0.5431, 0.5445)$ using either $\hat{K}$ or $\hat{D}$ (as $\hat{K}$ and $\hat{D}$ only differ by a constant under permutation of the orientations and thus lead to equivalent tests).

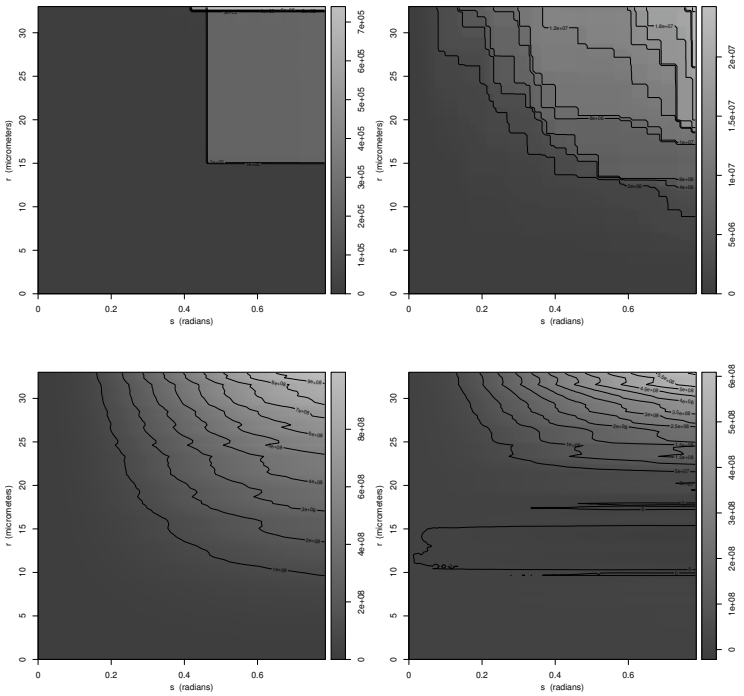

Fig. A.7: For $\hat{T} = \hat{K}$ (first row) and $\hat{T} = \hat{D}$ (second row): difference between $\hat{T}(r,s)$ for the observed neuron locations and orientations and the lower ($\hat{T}_{\text{low}}$) or upper ($\hat{T}_{\text{upp}}$) 95% global rank envelope under the fitted inhomogeneous Poisson model on $W \times \mathbb{S}^2$. Left: $\hat{T}(r,s) - \hat{T}_{\text{low}}(r,s)$. Right: $\hat{T}_{\text{upp}}(r,s) - \hat{T}(r,s)$.

6.3 Simulation study

In the data analyses in Sections 6.1–6.2, the tests based on $\hat{K}$ failed to reject the proposed Poisson models in cases where the corresponding spatial model was rejected when using $\hat{K}_1$. To investigate whether the space-sphere K-function is a valuable addition to the existing functional summary statistics on the space and sphere, we performed a simulation study comparing the power of global rank envelope tests based on either $\hat{K}$, $\hat{D}$, or a combination of $\hat{K}_1$ and $\hat{K}_2$. (The combined test function is simply a concatenation of $\hat{K}_1$ and $\hat{K}_2$. Mrkvička et al., 2017, recommended using such a combination rather than $\hat{K}_1$ or $\hat{K}_2$ as a test function.)

Specifically, we consider a homogeneous LGCP X driven by a random field $\Lambda(y,u) = \rho \exp\{Z(y,u)\}$, where $\rho > 0$ and

$$Z(y,u) = \alpha + \sigma_1 Z_1(y) + \sigma_2 Z_2(u) + \delta Z_3(y,u), \qquad y \in \mathbb{R},\ u \in \mathbb{S}^2,$$

for parameters $\sigma_1, \sigma_2 > 0$, $\delta \geq 0$, and $\alpha = -(\sigma_1^2 + \sigma_2^2 + \delta^2)/2$. Further,

Z_1, Z_2, and Z_3 are independent GRFs with mean 0 and covariance functions $c_1(y_1, y_2) = \exp(-\|y_1 - y_2\|_d/\phi_1)$, $c_2(u_1, u_2) = \exp(-d(u_1, u_2)/\phi_2)$, and $c_3(y_1, u_1, y_2, u_2) = c_1(y_1, y_2)c_2(u_1, u_2)$, respectively, with parameters $\phi_1, \phi_2 > 0$. Note that the resulting LGCP is homogeneous (and thus first order separable) and SOIRS for any value of $\delta \geq 0$. In addition, by (A.12), the process is second order separable if and only if $\delta = 0$, in which case X has pair correlation function

$$\begin{aligned} &g_\theta(y_1, u_1, y_2, u_2) \\ &\quad = \exp\{\sigma_1^2 c_1(y_1, y_2) + \sigma_2^2 c_2(u_1, u_2)\}, \qquad y_1, y_2 \in \mathbb{R},\ u_1, u_2 \in \mathbb{S}^2, \end{aligned} \tag{A.23}$$

where $\theta = (\sigma_1, \phi_1, \sigma_2, \phi_2)$.

For each value of $\delta = 0, 0.5, 1, 1.5, 2$, we simulated 100 realisations of a LGCP on $[0,1] \times \mathbb{S}^2$ with $\rho = 1000$, $\sigma_1 = \sigma_2 = 0.5$, $\phi_1 = 0.05$, and $\phi_2 = 0.132$. Then for each of these simulations, we fitted the LGCP model with $\delta = 0$ using a second order composite likelihood approach proposed by Guan (2006) to estimate θ. In the present time-sphere setting, for a finite point pattern $x \subset [0,1] \times \mathbb{S}^2$, the log second order composite likelihood is given by

$$\begin{aligned} \mathrm{CL}(\theta; x) = &\sum_{(y_i,u_i),(y_j,u_j) \in x}^{\neq} w(y_i, u_i, y_j, u_j) \log\{\rho_\theta^{(2)}(y_i, u_i, y_j, u_j)\} \\ &- n_{r,s} \log\Big\{ \int_{[0,1]\times\mathbb{S}^2} \int_{[0,1]\times\mathbb{S}^2} w(y_1, u_1, y_2, u_2) \\ &\qquad \cdot \rho_\theta^{(2)}(y_1, u_1, y_2, u_2)\, \mathrm{d}\mu(y_1, u_1)\, \mathrm{d}\mu(y_2, u_2) \Big\}. \end{aligned} \tag{A.24}$$

Here, for user specified distances r and s, $w(y_1, u_1, y_2, u_2) = \mathbb{I}\{\|y_1 - y_2\|_d < r,\ d(u_1, u_2) < s\}$, $n_{r,s}$ is the number of (r,s)-close neighbours, and $\rho_\theta^{(2)}$ is the second order joint intensity function, which for the homogeneous LGCP presented above is $\rho_\theta^{(2)}(y_1, u_1, y_2, u_2) = \rho^2 g_\theta(y_1, u_1, y_2, u_2)$. Then (A.24) is easily seen not to depend on ρ, and by (A.23) the composite likelihood can be written as

$$\mathrm{CL}(\theta; x) = l_1(\sigma_1, \phi_1; x) + l_2(\sigma_2, \phi_2; x) \tag{A.25}$$

for functions l_1 and l_2. Thus, maximising the composite likelihood with respect to θ can be split into two maximisation problems; that is, maximising l_1 with respect to (σ_1, ϕ_1) and l_2 with respect to (σ_2, ϕ_2). Finally, we tested the null hypothesis $\delta = 0$ using the global rank envelope test with 4999 simulations from the fitted model using either $\hat{K}$, $\hat{D}$, or a combination of $\hat{K}_1$ and $\hat{K}_2$ as test functions.

Table A.1 gives an overview of the conclusions reached by these tests. Note that the power of the tests based on either of the three test functions in

	Test function	$\delta = 0$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$
Liberal	$\hat{K}$	4%	7%	42%	75%	98%
	$\hat{D}$	2%	45%	92%	97%	100%
	$\hat{K}_1, \hat{K}_2$	10%	11%	29%	28%	42%
Conservative	$\hat{K}$	2%	5%	32%	72%	90%
	$\hat{D}$	0%	26%	77%	82%	86%
	$\hat{K}_1, \hat{K}_2$	10%	11%	29%	28%	40%

Table A.1: Power of tests for different values of δ when using the global rank envelope test with either $\hat{K}$, $\hat{D}$, or $\hat{K}_1$ combined with $\hat{K}_2$. The decision was made using a significance level of 5% for both the liberal and conservative tests

general increases with δ, both for the liberal and conservative test (for details see Appendix A). Thus, with increasing degree of non-separability the tests more often detect deviation from the separable model. However, tests based on $\hat{K}$ and particularly $\hat{D}$ seem preferable in this setup as they have a higher power than tests based on $\hat{K}_1$ combined with $\hat{K}_2$.

Obviously, the conservative p-value always lead to fewer rejections than the liberal, giving a lower power. However, if the global rank envelope procedure is based on a higher number of simulations, then the conservative and liberal test will more often lead to the same conclusion.

7 Additional comments

Section 2.2 introduced examples of space-sphere point processes for which the second order separability property described in Section 4 seems natural. However, for other classes of point processes a different structure of the pair correlation function may be more interesting. For example, suppose X is a Cox process driven by

$$\Lambda(y,u) = \sum_{(y',u',\gamma')\in\Phi} \gamma' k(y',u',y,u), \qquad y \in \mathbb{R}^d,\, u \in \mathbb{S}^k, \tag{A.26}$$

where Φ is a Poisson process on $S \times (0,\infty)$ with intensity function ζ, and $k(y',u',\cdot,\cdot)$ is a density with respect to μ. Then X is called a *shot noise Cox process (SNCP)* with kernel k (Møller, 2003). The process has intensity function

$$\rho(y,u) = \iint \gamma'\zeta(y',u',\gamma')k(y',u',y,u)\,\mathrm{d}\mu(y',u')\,\mathrm{d}\gamma', \qquad y \in \mathbb{R}^d,\, u \in \mathbb{S}^k,$$

and pair correlation function

$$g(y_1, u_1, y_2, u_2) = 1 + \frac{\iint \gamma'^2 \zeta(y', u', \gamma') k(y', u', y_1, u_1) k(y', u', y_2, u_2) \, \mathrm{d}\mu(y', u') \, \mathrm{d}\gamma'}{\rho(y_1, u_1)\rho(y_2, u_2)} \quad \text{(A.27)}$$

for any $y_1, y_2 \in \mathbb{R}^d$ and any $u_1, u_2 \in \mathbb{S}^k$ with $\rho(y_1, u_1)\rho(y_2, u_2) > 0$. In the trivial case where the kernel $k(y', u', y, u)$ in (A.26) does not depend on u (or y), the SNCP is both first and second order separable, with intensity and pair correlation functions that do not depend on the spherical (or spatial) components, and the process thus fulfils second order separability in the sense of (A.17). However, the specific structure of the pair correlation function for a SNCP in (A.27) makes it more natural to look for a product structure in $g - 1$ rather than g. That is, we may say that X is second order separable if there exist Borel functions h_1 and h_2 such that

$$g(y_1, u_1, y_2, u_2) - 1 = h_1(y_1, y_2) h_2(u_1, u_2) \quad \text{(A.28)}$$

for $y_1, y_2 \in \mathbb{R}^d$, $u_1, u_2 \in \mathbb{S}^k$. This property is naturally fulfilled whenever we consider a Poisson process or any marked point process with marks that are IID and independent of the ground process as described in Example 3. Now, think of Φ in (A.26) as a marked point process with ground process $\{(y, u) : (y, u, \gamma) \in \Phi\}$ and marks $\{\gamma : (y, u, \gamma) \in \Phi\}$, and assume that the ground process and the marks are independent processes, the ground process is a homogeneous Poisson process on S with intensity $\alpha > 0$, and the marks are IID with mean m_1 and second moment m_2. If in addition $k(y', u', y, u) = k_0\{y - y', d(u, u')\}$, then Λ and thus X is stationary in space and isotropic on the sphere. Further, X is homogeneous with intensity $\rho = m_1\alpha$ and pair correlation function

$$g(y_1, u_1, y_2, u_2) = 1 + \frac{m_2}{\alpha m_1^2} \int k_0\{y_1 - y', d(u_1, u')\} k_0\{y_2 - y', d(u_2, u')\} \, \mathrm{d}\mu(y', u') \quad \text{(A.29)}$$

for $y_1, y_2 \in \mathbb{R}^d$ and $u_1, u_2 \in \mathbb{S}^k$. Clearly, (A.29) depends only on (y_1, y_2) through $y_1 - y_2$, and on (u_1, u_2) through $d(u_1, u_2)$, although there is no simple expression for these dependencies in general. Furthermore, separability in the form of (A.28) is fulfilled if the kernel k in (A.29) factorizes such that

$$k_0\{y - y', d(u, u')\} = k_{01}(y - y') k_{02}\{d(u, u')\}, \qquad y, y' \in \mathbb{R}^d, \, u, u' \in \mathbb{S}^k,$$

for Borel functions k_{01} and k_{02}. Then by (A.6)–(A.9), the pair correlations functions for Y and U_W are

$$g_1(y_1, y_2) = 1 + c_1 \frac{m_2}{\alpha m_1^2} \int k_{01}(y_1 - y') k_{01}(y_2 - y') \, \mathrm{d}y', \qquad y_1, y_2 \in \mathbb{R}^d,$$

and

$$g_2(u_1, u_2) = 1 + c_2 \frac{m_2}{\alpha m_1^2} \int k_{02}\{d(u_1, u')\} k_{02}\{d(u_2, u')\} \, \mathrm{d}\nu(u'), \qquad u_1, u_2 \in \mathbb{S}^k,$$

where

$$c_1 = \frac{1}{\sigma_k^2} \iiint k_{02}\{d(u_1, u')\} k_{02}\{d(u_1, u')\} \, \mathrm{d}\nu(u_1) \, \mathrm{d}\nu(u_2) \, \mathrm{d}\nu(u')$$

and

$$c_2 = \frac{1}{|W|^2} \int \int_W \int_W k_{01}(y_1 - y') k_{01}(y_1 - y') \, \mathrm{d}y_1 \, \mathrm{d}y_2 \, \mathrm{d}y'.$$

That is, g_1 depends only on the spherical components through the constant c_1, and similarly g_2 depends only on the spatial components through c_2. Prokešová and Dvořák (2014) discuss how an analogue property can be exploited to estimate the parameters of space-time SNCPs using minimum contrast estimation for the projected processes. A similar procedure will be applicable for space-sphere point processes, but we have not investigated this further.

Section 5 considered the situation where the spatial components of X are observed on a subset of $\mathbb{R}^d$ and the spherical components are observable on the entire sphere. In more general applications, the spherical components may only be observable on a subset of $\mathbb{S}^k$ leading to edge effects on the sphere too. To account for this, edge correction methods for the sphere should be used when estimating K_2 (see Lawrence et al., 2016) and K. If X is observable on a product space $W_1 \times W_2$, where $W_1 \subset \mathbb{R}^d$ and $W_2 \subset \mathbb{S}^k$, then an edge corrected estimate for K may be obtained by combining edge corrected estimates for K_1 and K_2 analogous to (A.22). Concerning the specific choice of edge correction method, Baddeley et al. (2015) mentioned for planar point processes that, "So long as some kind of edge correction is performed ..., the particular choice of edge correction technique is usually not critical". We expect that the situation is similar for our setting.

For one-dimensional test functions, Myllymäki et al. (2017) recommend using 2499 simulations to perform a global rank envelope test, and Mrkvička et al. (2017) discuss the appropriate number of simulations when using a multivariate test function (as the empirical space-sphere K-function). In Section 6.2, we used 49999 simulations for the global rank envelope test based on $\hat{K}$, since $\hat{K}$ had steep jumps. To avoid this large number of simulations, a refinement of the global rank envelope test discussed in Mrkvička et al. (2018) can be applied.

In Example 1 we noticed that if a space-sphere point process is a Poisson process, then the spatial and spherical components are Poisson processes as well. Nevertheless, using $\hat{K}$ for testing a space-sphere Poisson model may

lead to a different conclusion than using $\hat{K}_1$ and $\hat{K}_2$ for testing whether the corresponding Poisson models for the spatial and spherical components, respectively, are appropriate. Indeed, in the case of Figures A.3–A.4, the test based on $\hat{K}_1$ showed some evidence against a homogeneous Poisson model for the fireball event times, while no evidence against a homogeneous Poisson model for the locations over time was seen with the test based on $\hat{K}$. This observation together with the results in Section 6.2 was our motivation for making the simulation study in Section 6.3, where we investigated the power of global rank envelope tests based on either $\hat{K}$, $\hat{D}$, or a combination of $\hat{K}_1$ and $\hat{K}_2$, and we concluded that tests based on $\hat{K}$ and in particular $\hat{D}$ seem preferable.

In Section 6.3, we utilised homogeneity and second order separability to speed up optimisation of the second order composite likelihood proposed in Guan (2006). If the process is inhomogeneous but first (and second) order separable, we still get a separation of the composite likelihood similar to (A.25), where l_1 and l_2 now may depend also on intensity parameters. As an alternative, the second order composite likelihood discussed in Waagepetersen (2007) can be used. However, in that case, first and second order separability do not yield a separable likelihood as in (A.25), and for our simulation study it resulted in unstable estimates (and thus it was discarded in favour of the one proposed by Guan, 2006). Furthermore, one may investigate whether the adaptive procedure discussed in Lavancier et al. (2018) will provide stable estimates in the space-sphere setting. In short, Lavancier et al. (2018) consider the score function related to (A.24) and introduce a modified weight function w depending on g.

In this paper, we considered point processes living on $\mathbb{R}^d \times \mathbb{S}^k$. Naturally, we may extend the results/methods to more general metric spaces $\mathbb{R}^d \times \mathbb{M}$, where $\mathbb{M}$ is a compact set (e.g. a torus). However, we need to require some invariance property for the metric space $\mathbb{M}$ and its metric under a group action, such that we can define an equivalence of the SOIRS property needed to define K.

Acknowledgements. The authors are grateful to Jiří Dvořák for helpful comments and to Ali H. Rafati for collecting the pyramidal cell data. This work was supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 "Statistics for point processes in space and beyond", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation.

Appendix A

In Sections 6.1–6.3, we used the global rank envelope test presented in Myllymäki et al. (2017) to test for various point process models. In this appendix, we briefly explain the idea and use of such a test. A global rank envelope test compares a chosen test function for the observed data with the distribution of the test function under the null model; as this distribution is typically unknown it is approximated using a Monte Carlo approach. The comparison is based on a rank that only gives a weak ordering of the test functions. Thus, instead of a single p-value, the global rank envelope test provides an interval of p-values, where the end points specify the most liberal and conservative p-values of the test. A narrow p-interval is desirable as the test is inconclusive if the p-interval contains the chosen significance level. The width of the p-interval depends on the number of simulations, smoothness of the test functions and dimensionality. Myllymäki et al. (2017) recommended to use 2499 simulations for one-dimensional test functions and a significance level of 5%.

An advantage of the global rank envelope procedure is that it provides a graphical interpretation of the test in form of critical bounds (called a global rank envelope) for the test function. For example, if the observed test function is not completely inside the 95% global rank envelope, this corresponds to a rejection of the null hypothesis at a significance level of 5%. Furthermore, locations where the observed test function falls outside the global rank envelope reveal possible reasons for rejecting the null model.

In their supplementary material, Myllymäki et al. (2017) discussed two approaches for calculating test functions that rely on an estimate of the intensity. One approach is to reuse the intensity estimate for the observed point pattern in calculation of all the test functions, another is to reestimate the intensity for each simulation and then use this estimate when calculating the associated test function. For the L-function, which is a transformation of K_1, Myllymäki et al. (2017) concluded that the reestimation approach give the more powerful test. In this paper, we have therefore based all our global rank envelope tests on that approach.

References

Baddeley, A., Møller, J., and Waagepetersen, R. P. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350.

Baddeley, A., Nair, G., Rakshit, S., and McSwiggan, G. (2017). "Stationary" point processes are uncommon on linear networks. *Stat*, 6:68–78.

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Buxhoeveden, D. P. and Casanova, M. F. (2002). The minicolumn hypothesis in neuroscience. *Brain*, 125:935–951.

Cox, D. R. (1955). Some statistical models related with series of events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 17:129–164.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer-Verlag, New York, second edition.

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Diggle, P. J., Chetwynd, A., Häggkvist, R., and Morris, S. E. (1995). Second-order analysis of space-time clustering. *Statistical Methods in Medical Research*, 4:124–136.

Dvořák, J. and Prokešová, M. (2016). Parameter estimation for inhomogeneous space-time shot-noise Cox point processes. *Scandinavian Journal of Statistics*, 43:939–961.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge University Press, New York.

Gabriel, E. and Diggle, P. J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63:43–51.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101:1502–1512.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Statistics in Practice. Wiley, New York.

Koubek, A., Pawlas, Z., Brereton, T., Kriesche, B., and Schmidt, V. (2016). Testing the random field model hypothesis for random marked closed sets. *Spatial Statistics*, 16:118–136.

Lavancier, F., Møller, J., and Rubak, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77:853–877.

Lavancier, F., Poinas, A., and Waagepetersen, R. P. (2018). Adaptive estimating function inference for non-stationary determinantal point processes. Available on arXiv:1806.06231.

Lawrence, T., Baddeley, A., Milne, R., and Nair, G. (2016). Point pattern analysis on a region of a sphere. *Stat*, 5:144–157.

Li, S. (2011). Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics & Statistics*, 4:66–70.

Lorente de Nó, R. (1938). The cerebral cortex: Architecture, intracortical connections, motor projections. In Fulton, J. F., editor, *Physiology of the Nervous System*, pages 274–301. Oxford University Press, Oxford.

Møller, J. and Rubak, E. (2016). Functional summary statistics for point processes on the sphere with an application to determinantal point processes. *Spatial Statistics*, 18:4–23.

Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability*, 35:614–640.

Møller, J. and Ghorbani, M. (2012). Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. *Statistica Neerlandica*, 66:472–491.

Møller, J., Nielsen, M., Porcu, E., and Rubak, E. (2018). Determinantal point process models on the sphere. *Bernoulli*, 24:1171–1201.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25:451–482.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, Boca Raton, Florida.

Møller, J. and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34:643–684.

Mountcastle, V. B. (1978). *The Mindful Brain: Cortical Organization and the Group-selective Theory of Higher Brain Function*. MIT Press, Cambridge.

Mrkvička, T., Myllymäki, M., and Hahn, U. (2017). Multiple Monte Carlo testing, with applications in spatial point processes. *Statistics and Computing*, 27:1239–1255.

Mrkvička, T., Hahn, U., and Myllymäki, M. (2018). A one-way ANOVA test for functional data with graphical interpretation. Available on arXiv:1612.03608.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Mathematische Operationsforschung und Statistik, series Statistics*, 14:63–71.

Prokešová, M. and Dvořák, J. (2014). Statistics for inhomogeneous space-time shot-noise Cox processes. *Spatial Statistics*, 10:76–86.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258.

Paper B

Pair correlation functions and limiting distributions of iterated cluster point processes

Jesper Møller and Andreas D. Christoffersen

The paper has been published in the
Journal of Applied Probability Vol. 55, pp. 789–809, 2018.

The layout has been revised.

Abstract

We consider a Markov chain of point processes such that each state is a superposition of an independent cluster process with the previous state as its centre process together with some independent noise process and a thinned version of the previous state. The model extends earlier work by Felsenstein and Shimatani describing a reproducing population. We discuss when closed term expressions of the first and second order moments are available for a given state. In a special case it is known that the pair correlation function for these type of point processes converges as the Markov chain progresses, but it has not been shown whether the Markov chain has an equilibrium distribution with this, particular, pair correlation function and how it may be constructed. Assuming the same reproducing system, we construct an equilibrium distribution by a coupling argument.

1 Introduction

This paper deals with a discrete time Markov chain of point processes $G_0, G_1, \ldots$ in the d-dimensional Euclidean space $\mathbb{R}^d$, where the chain describes a reproducing population and we refer to G_n as the nth generation (of points). We make the following assumptions. Any point process considered in this paper will be viewed as a random subsets of $\mathbb{R}^d$ which is almost surely locally finite, that is, the point process has almost surely a finite number of points within any bounded subset of $\mathbb{R}^d$ (for measure theoretical details, see e.g. Daley and Vere-Jones (2003) or Møller and Waagepetersen (2004)). Recall that a point process $X \subset \mathbb{R}^d$ is stationary if its distribution is invariant under translations in $\mathbb{R}^d$, and then its intensity $\rho_X \in [0, \infty]$ is the mean number of points in X falling in any Borel subset of $\mathbb{R}^d$ of unit volume. Now, for generation 0, G_0 is stationary with intensity $\rho_{G_0} \in (0, \infty)$. Further, for generation $n = 1, 2, \ldots$, conditional on the previous generations $G_0, \ldots, G_{n-1}$, we obtain G_n by four basic operations for point processes:

(a) Independent clustering: To each point $x \in G_{n-1}$ is associated a (non-centred) cluster $Y_{n,x} \subset \mathbb{R}^d$. These clusters are independent identically distributed (IID) finite point processes and they are independent of $G_0, \ldots, G_{n-1}$. The cardinality of $Y_{n,x}$ has finite mean β_n and finite variance ν_n and is independent of the points in $Y_{n,x}$ which are IID, with each point following a probability density function (PDF) f_n. We refer to $x + Y_{n,x}$ (the translation of $Y_{n,x}$ by x) as the offspring/children process generated by the ancestor/parent x, and we let

$$Y_n = \bigcup_{x \in G_{n-1}} (x + Y_{n,x}) \tag{B.1}$$

be the independent cluster process given by the superposition of all offspring processes generated by the points in the previous generation G_{n-1}.

(b) Independent thinning: For all $y \in \mathbb{R}^d$, let $B_{n,y}$ be IID Bernoulli variables which are independent of Y_n, $G_0, \ldots, G_{n-1}$, and all previously generated Bernoulli variables. Let $p_n = \mathrm{P}(B_{n,y} = 1)$. For all $x \in G_{n-1}$, let

$$W_{n,x} = \{y \in x + Y_{n,x} : B_{n,y} = 1\}$$

be the independent p_n-thinned point process of $x + Y_{n,x}$, and let

$$W_n = \bigcup_{x \in G_{n-1}} W_{n,x} \tag{B.2}$$

be the independent p_n-thinned point process of Y_n. Note that with probability one, $W_n \cap G_{n-1} = \emptyset$, since by assumption on the cluster points the origin is not contained in $Y_{n,x}$.

(c) Independent retention: For all $x \in \mathbb{R}^d$, let $Q_{n,x}$ be IID Bernoulli variables which are independent of Y_n, $G_0, \ldots, G_{n-1}$, and all previously generated Bernoulli variables. Let $q_n = P(Q_{n,x} = 1)$ and let

$$G_{n-1}^{\text{thin}} = \{x \in G_{n-1} : Q_{n,x} = 1\}$$

be the independent q_n-thinned point process of G_{n-1}.

(d) Independent noise: Let $Z_n \subset \mathbb{R}^d$ be a stationary point process with finite intensity ρ_{Z_n} and independent of W_n, $G_0, \ldots, G_{n-1}$, and G_{n-1}^{thin}. Finally, let

$$G_n = W_n \cup G_{n-1}^{\text{thin}} \cup Z_n \tag{B.3}$$

where we interpret Z_n as noise. For ease of presentation we assume with probability one that $W_n \cup G_{n-1}^{\text{thin}}$ and Z_n are disjoint. Thus W_n, G_{n-1}^{thin}, and Z_n are pairwise disjoint almost surely.

When we later interpret our results, for any point $x \in G_{n-1}^{\text{thin}}$, since $x \in G_{n-1} \cap G_n$, we consider x both as its own ancestor and its own child.

Our model is an extension of the model in Shimatani's paper Shimatani (2010), which in turn is an extension of Malécot's model studied in Felsenstein (1975) (we return to this in Section 2, item (vii) and (viii)). In particular, our extension allows us to model cluster centres exhibiting clustering or regularity, points from previous generations can be retaining, and the noise processes can also exhibit clustering or regularity (i.e., they are not assumed to be Poisson processes). For statistical applications, we have in mind that G_n may be observable (at least for some values of $n \geq 1$) whilst G_0 and the cluster, thinning, and superpositioning procedures in items (a)–(b) and (d) are

unobservable. Our model may be of relevance for applications in population genetics and community ecology (see Shimatani (2010) and the references therein), for analyzing tropical rain forest point pattern data with multiple scales of clustering (see Wiegand et al. (2007)), and for modelling proteins with multiple noisy appearances in PhotoActivated Localization Microscopy (PALM) (see Andersen et al. (2018)). However, we leave it for other work to study the statistical applications of our model and results.

The paper is organized as follows. A discussion of the assumptions in items (a)–(d) and the related literature is given in Section 2. Section 3 focuses on the first and second order moment properties of G_n, that is, its intensity and pair correlation function (PCF); we extend model cases and results in Shimatani's paper Shimatani (2010) and show that tractable model cases for the PCF of G_0 are meaningful in terms of Poisson and other point processes, including weighted permanental and weighted determinantal point processes (which was not observed in Shimatani (2010)). Section 4 discusses limiting cases of the PCF of G_n as $n \to \infty$ when we have the same reproduction system and under weaker conditions than in Shimatani (2010). In particular, when natural conditions are satisfied, we establish ergodicity of the Markov chain by using a coupling construction and by giving a constructive description of the chain's unique invariant distribution when extending the Markov chain backwards in time. Finally, Appendix A provides background knowledge on weighted permanental and determinantal point processes, Appendix B verifies some technical details, and Appendix C specifies an algorithm for approximate simulation of the Markov chain's invariant distribution.

2 Assumptions and related work

Items (i)–(iv) below comment on the model assumptions in items (a)–(d).

(i) The process Y_n is a stationary independent cluster process Daley and Vere-Jones (2003) and we have the following special cases: If G_{n-1} is a stationary Poisson process, Y_n is a Neyman-Scott process Neyman and Scott (1958); if in addition $\#Y_{n,x}$ follows a Poisson distribution, then $\beta_n = \nu_n$ and Y_n is a shot-noise Cox process (SNCP; Møller (2003)) driven by

$$\Lambda_n(x) = \beta_n \sum_{y \in G_{n-1}} f_n(x-y), \qquad x \in \mathbb{R}^d. \tag{B.4}$$

This is a (modified) Thomas process Thomas (1949) if f_n is the density of d IID zero-mean normally distributed variates with variance σ_n^2 – we denote this distribution by $N_d(\sigma_n^2)$ – and it is a Matérn cluster process Matérn (1960, 1986) if instead f_n is a uniform density of a d-dimensional ball with centre at the origin. However, in many applications a Poisson

centre process is not appropriate. For instance, Van Lieshout & Baddeley (2002) considered a repulsive Markov point process model for the centre process, whereby it is easier to identify the clusters than under a Poisson centre process.

(ii) When $\beta_n \leq \nu_n$, we may consider Y_n as a stationary generalised shot-noise Cox process (GSNCP; see Møller and Torrisi (2005)). In this model (B.4) is extended to the case where G_{n-1} is a general stationary point process and Y_n is a Cox process driven by

$$\Lambda_n(x) = \sum_{y \in G_{n-1}} \gamma_y k_n[\{(x-y)/b_y\}]/b_y^d, \qquad x \in \mathbb{R}^d,$$

where k_n is a PDF on $\mathbb{R}^d$, the γ_y and the b_y for all $y \in G_{n-1}$ are independent positive random variables which are independent of G_{n-1}, and the γ_y are identically distributed with mean β_n and variance $\nu_n - \beta_n$ (as $\#Y_{n,x}$ has mean β_n and variance $\nu_n = \mathrm{E}\{\mathrm{var}(\#Y_{n,x}|\gamma_y)\} + \mathrm{var}\{\mathrm{E}(\#Y_{n,x}|\gamma_y)\}$ $= \beta_n + \mathrm{var}(\gamma_n)$). Further, b_y has an interpretation as a random bandwidth and

$$f_n(x) = \mathrm{E}\left\{\frac{k_n(x/b_y)}{b_y^d}\right\}.$$

The general results for the intensity and PCF of G_n in Section 3 will be unchanged whether we consider this stationary GSNCP or the more general case in item (a).

(iii) Clearly, there is no noise (Z_n is empty with probability one) if $\rho_{Z_n} = 0$. The case $\rho_{Z_n} > 0$ may be relevant when not all points in a generation can be described as resulting from independent clustering and thinning as in (a)–(c). Note that in item (d) we could without loss of generality assume $Z_1, Z_2, \ldots$ are independent. Further, we introduce the thinning of Y_n in item (b) only for modelling purposes and for comparison with Shimatani (2010); from a mathematical point of view this thinning could be omitted if in item (a) we replace each cluster $Y_{n,x}$ by what happens after the independent thinning: Namely that independent thinned clusters $Y_{n,x}^{\mathrm{th}}$ appear so that $\#Y_{n,x}^{\mathrm{th}}$ has mean $\beta_n^{\mathrm{th}} = \beta_n p_n$ and variance $\nu_n^{\mathrm{th}} = \beta_n p_n - \beta_n p_n^2 + \nu_n p_n^2$ and is independent of the points in $Y_{n,x}^{\mathrm{th}}$ which are IID with PDF f_n, whereby W_n and $Y_n^{\mathrm{th}} := \cup_{x \in G_{n-1}}(x + Y_{n,x}^{\mathrm{th}})$ are identically distributed.

(iv) Assuming for $n = 1, 2, \ldots$ no thinning of Y_n ($p_n = 1$), an equivalent description of items (a) and (c)–(d) is given in terms of the Voronoi tessellation generated by G_{n-1}: For $x \in G_{n-1}$, let $C(x|G_{n-1})$ be the Voronoi cell associated to x and consisting of all points in $\mathbb{R}^d$ which are at least as close to x than to any other point in G_{n-1} (with respect to

usual distance in $\mathbb{R}^d$). With probability one, since G_{n-1} is stationary and non-empty, each Voronoi cell is bounded and hence its volume is finite (see e.g. Møller (1989, 1994)). Thus we can set

$$G_n = \bigcup_{x \in G_{n-1}} (x + G_{n,x}) \bigcup G_{n-1}^{\text{thin}}$$

where conditional on G_{n-1} and for all $x \in G_{n-1}$, the $G_{n,x}$ are independent of G_{n-1}^{thin} and they are IID finite point processes with a distribution as follows: $\#G_{n,x}$ has mean $\beta_n + |C(x|G_{n-1})|\rho_{Z_n}$, variance $\nu_n + |C(x|G_{n-1})|\rho_{Z_n}$, and is independent of the points in $G_{n,x}$, where $|\cdot|$ denotes volume. The points in $G_{n,x}$ are i.i.d., each following a mixture distribution so that with probability $\beta_n / \{\beta_n + |C(x|G_{n-1})|\rho_{Z_n}\}$ the PDF is f_n and else it is a uniform distribution on $C(x|G_{n-1})$.

In items (v)–(vi) below we discuss earlier work on the model for $G_0, G_1, \ldots$, where G_0 is a stationary Poisson process, all $G_n = Y_n$ for $n \geq 1$ (i.e., no thinning, no retention, and no noise), $f_n = f$ and $\beta_n = \beta$ do not depend on $n \geq 1$. We may refer to this as a replicated SNCP. Frequently in the literature, a so-called replicated Thomas process is considered, that is, $f \sim N_d(\sigma^2)$.

(v) Apparently this replicated SNCP was originally studied by Malécot, see the discussion and references in Felsenstein's paper Felsenstein (1975) where the following three conditions are stated:

  (I) "individuals are distributed randomly on the line with equal expected density everywhere";

  (II) "each individual reproduces independently, the number of offspring being drawn from a Poisson distribution with a mean of one"; and

  (III) "each offspring migrates independently, the displacements being drawn from some distribution m(x), which we will take to be a normal distribution."

  (In our notation, $d = 1$, $\beta = 1$, and $f \sim N_1(\sigma^2)$, but Felsenstein (1975) considered also more general offspring densities f and the cases $d = 2, 3$.) Felsenstein (1975) noted that "(I) is incompatible with (II)–(III)" because $G_1, G_2, \ldots$ are not stationary Poisson processes and "a model embodying (II) and (III) will lead to the formation of larger and larger clumps of individuals separated by greater and greater distances", and then he concluded "This model is therefore biologically irrelevant".

(vi) Kingman in Kingman (1977) considered the case where β is replaced by a non-negative function b which is allowed to depend on the cluster centre x and the previous generation, so a cluster with centre x is a

Poisson process with intensity function $b(x, G_{n-1})f(\cdot - x)$; e.g., as in the Voronoi case discussed in item (iv), $b(x, G_{n-1})$ may depend on G_{n-1} in a neighbourhood of x. Then G_n is a Cox process: G_n conditional on G_{n-1} is a Poisson process with intensity function

$$\Lambda_n(x) = \sum_{y \in G_{n-1}} b(y, G_{n-1}) f(x - y), \qquad x \in \mathbb{R}^d. \tag{B.5}$$

In this setting Kingman (1977) verified that it is impossible for G_n to be a stationary Poisson process, however, replacing $f(x-y)$ in (B.5) by a more general density which may depend on $G_{n-1} - x$, Kingman (1977) noticed that it is possible for G_n to be a stationary Poisson process. A trivial example is the Voronoi case in item (iv) when $G_n = Z_n$ for $n \geq 1$.

Recently, Shimantani in Shimatani (2010) considered first the case of items (a)–(b) and no noise, when $d = 2$ and there is the same reproduction system so that $f_n = f$, $\beta_n = \beta > 0$, $\nu_n = \nu$, and $p_n = p \in (0, 1]$ do not depend on $n \geq 1$.

(vii) In particular, Shimatani (2010) considered the case $f \sim N_2(\sigma^2)$ and when $\beta p = 1$ or equivalently when the intensities $\rho_{G_0} = \rho_{G_1} = \ldots$ are invariant over generations, and then he showed that as $n \to \infty$, the PCF for G_n diverges. It follows from item (iii) that the model is equivalent to a replicated Neyman-Scott process; this becomes a replicated Thomas process when each cluster size is Poisson distributed, and hence the result in Shimatani (2010) agrees with the results in Felsenstein (1975) and Kingman (1977). Note that Shimatani (2010) implicitly assumed that a cluster can have more than one point. Otherwise the PCF of G_n becomes equal to 1; we discuss this rather trivial case again in Section 3.2 and 4; see also Section 3 in Kingman (1977).

Then, Shimantani in Shimatani (2010) extended the model by including noise as in item (d) and by making the following assumptions: The noise processes Z_n are stationary Poisson processes, satisfying $0 < \rho_{Z_1} = \rho_{Z_2} = \ldots$ and $\rho_{G_0} = \rho_{G_1} = \ldots$, meaning that $\beta p \leq 1$. As there is no noise if $\beta p = 1$ it is also assumed that $\beta p < 1$.

(viii) Then Shimatani (2010) showed that the PCF of G_n converges uniformly as $n \to \infty$ and he argued that this limiting case may be "biologically valid" (Shimatani, 2010, Section 2.4). However, we address some points arising from Shimatani (2010).

- He did not show that there exists an underlying point process having this limiting case as its PCF, although he claimed that "this modified replicated Neyman-Scott process reaches an equilibrium

state". In Section 4, for our more general model, we prove the existence of such an underlying point process.

- When G_0 is not a stationary Poisson process but its PCF is of a particular form (which we specify later in connection to (B.9)), he did not argue that there exists an underlying point process and what it could be. In Section 3, we verify this existence under our more general model.

Finally, we remark on a few related cases.

(ix) Whilst we study the processes G_n for all $n = 1, 2, \ldots$, often in the spatial point process literature the focus is on either G_1 or G_2, assuming $p_n = 1$ and $\rho_{Z_n} = 0$ for $n = 1$ or $n = 1, 2$, respectively. Wiegand et al. (2007) studied this in the special case of a double Thomas cluster process G_2 when $d = 2$, i.e., when G_0 is a stationary Poisson process, (B.4) holds for both $G_1 = Y_1$ and $G_2 = Y_2$, and $f_n \sim N_2(\sigma_n^2)$ for $n = 1, 2$; see also Andersen et al. (2018) for more general functions f_n. Moreover, Wiegand et al. (2007) extended the double Thomas process to the case where $\rho_{Z_1} = 0$ and $\rho_{Z_2} > 0$; this type of model is also considered in Andersen et al. (2018). In any case, our general results for intensities and PCFs in Section 3 will cover all these cases.

(x) If for each generation we assume no thinning ($p_1 = p_2 = \ldots = 1$), no noise ($\rho_{Z_1} = \rho_{Z_2} = \ldots = 0$), no retention ($q_1 = q_2 = \ldots = 0$) as well as $\beta_1 = \beta_2 = \ldots$ and $f_1 = f_2 = \ldots$, then the superposition $\bigcup_{n=0}^{\infty} G_n$ is known as a spatial Hawkes process, see Møller and Torrisi (2007) and the references therein.

3 First and second order moment properties

In this section we determine the intensity and the PCF of G_n for $n = 1, 2, \ldots$, under more general assumptions than in Shimatani's paper Shimatani (2010). Specifically, points from one generation can be retained in the next generation, the noise is an arbitrary stationary point process (not necessarily a stationary Poisson process as in Shimatani (2010)), and we do not assume the same reproduction system.

3.1 Intensities

By induction G_n is seen to be stationary for $n = 0, 1, \ldots$ Its intensity is determined in the following proposition where for notational convenience we define $Z_0 = G_0$ so that $\rho_{Z_0} = \rho_{G_0}$.

Proposition 3.1 . For $n = 1, 2, \ldots$, we have that G_n is stationary with a positive and finite intensity given by

$$\rho_{G_n} = \rho_{G_{n-1}}(\beta_n p_n + q_n) + \rho_{Z_n} = \rho_{Z_n} + \sum_{i=0}^{n-1} \rho_{Z_i} \prod_{j=i+1}^{n} (\beta_j p_j + q_j). \tag{B.6}$$

Proof. Using induction for $n = 1, 2, \ldots$, the proposition follows immediately from items (a)–(d), where the term $\rho_{Z_i} \prod_{j=i+1}^{n}(\beta_j p_j + q_j)$ is the contribution to the intensity caused by the clusters with centres Z_i and after applying the two types of independent thinnings. □

3.2 Pair correlation functions

Preliminaries

Recall that a stationary point process $X \subset \mathbb{R}^d$ with intensity $\rho_X \in (0, \infty)$ has a translation invariant PCF (pair correlation function) $(u, v) \to g_X(u - v)$ with $(u, v) \in \mathbb{R}^d \times \mathbb{R}^d$ if for any bounded Borel function $h : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ with compact support,

$$\mathrm{E} \sum_{x_1, x_2 \in X:\, x_1 \neq x_2} h(x_1, x_2) = \rho_X^2 \iint h(x_1, x_2) g_X(x_1 - x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 < \infty. \tag{B.7}$$

Equivalently, for any bounded and disjoint Borel sets $A, B \subset \mathbb{R}^d$, denoting $N(A)$ the cardinality of $X \cap A$, the covariance between $N(A)$ and $N(B)$ exists and is given by

$$\mathrm{cov}\{N(A), N(B)\} = \rho_X^2 \int_A \int_B \{g_X(x_1 - x_2) - 1\}\, \mathrm{d}x_1\, \mathrm{d}x_2.$$

Some remarks are in order. Note that g_X is uniquely determined except for nullsets with respect to Lebesgue measure on $\mathbb{R}^d$, but we ignore such nullsets in the following. Thus the translation invariance of the PCF is implied by the stationarity of X. Our results below are presented in terms of the reduced PCF $g_X - 1$ rather than g_X, and $g_X = 1$ if X is a Poisson process. It is convenient when g_X is isotropic, meaning that there is a function $g_{X,o}$ so that for all $x \in \mathbb{R}^d$, $g_X(x) = g_{X,o}(\|x\|)$ depends only on x through $\|x\|$. With a slight abuse of terminology, we also refer to g_X and $g_{X,o}$ as PCFs.

For a PDF h on $\mathbb{R}^d$, let $\tilde{h}(x) = h(-x)$ and let

$$h * \tilde{h}(x_1 - x_2) = \int h(x_1 - y) h(x_2 - y)\, \mathrm{d}y \tag{B.8}$$

be the convolution of h and $\tilde{h}$. Note that if U and V are IID random variables with PDF h, then $U - V$ has PDF $h * \tilde{h}$. In the following section we consider the case

$$g_X - 1 = a\, h * \tilde{h} \tag{B.9}$$

for real constants a, where X in particular, may refer to the initial generation process, G_0, or the noise process, Z_n. This corresponds to X being a Poisson process if $a = 0$, a point process with positive association between its points (attractiveness, clustering, or clumping) if $a > 0$, and a point process with negative association between its points (repulsiveness or regularity) if $a < 0$. In Shimatani (2010), for the initial generation process G_0, Shimatani briefly discussed the special case of (B.9) when $h \sim N_2(\tau^2/2)$ (so $h * \tilde{h} \sim N_2(\tau^2)$) whilst the noise processses are stationary Poisson processes. However, if $a \neq 0$ he did not argue if an underlying point process with PCF g_X exists. Indeed, as detailed in Appendix A, there exist α-weighted determinantal point processes satisfying (B.9) if $\alpha = -1/a$ is a positive integer, and there exist Cox processes given by α-weighted permanental point processes satisfying (B.9) if $\alpha = 1/a$ is a positive half-integer. Additionally, h needs not to be Gaussian when dealing with weighted determinantal and permanental point processes; e.g. h may be the density of a normal-variance mixture distribution Barndorff-Nielsen et al. (1982). Also generalized shot-noise Cox processes Møller and Torrisi (2007) have PCFs of the form (B.9) with $a > 0$. Moreover, (B.9) holds for many other cases of point process models for X: If the Fourier transform $\mathcal{F}(g_X - 1)$ is well-defined and non-negative, if $h = \tilde{h}$, and if $a := \int (g_X - 1) \in (0, \infty)$, then (B.9) holds with

$$h = \mathcal{F}^{-1}\{\sqrt{\mathcal{F}(g_X - 1)}\}/\sqrt{a}$$

provided this inverse transform is well-defined. Extensions of B.9 are discussed in Section 3.2

We will need the following lemma in Section 3.2.

Lemma 3.2 . Suppose g_X is of the form (B.9). Then

$$\iint \{g_X(x_1 - x_2) - 1\} f(u - x_1) f(v - x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 = a h * \tilde{h} * f * \tilde{f}(u - v)$$

for any integrable real function f defined on $\mathbb{R}^d$ and for any $u, v \in \mathbb{R}^d$.

Proof. Follows from (B.8) and (B.9) using Fubini's theorem and the fact that the convolution operation is commutative and associative. □

First main result

This section concerns our first main result, Theorem 3.3, which is verified in Section (3.2). We use the following notation. Define

$$c_n = \mathrm{E}\{\#Y_{n,x}(\#Y_{n,x} - 1)\}/\beta_n^2 = (\nu_n + \beta_n^2 - \beta_n)/\beta_n^2 \qquad \text{if } \beta_n > 0, \tag{B.10}$$

with $c_n = 0$ if $\beta_n = 0$. If $\beta_n = \nu_n > 0$, as in the case when $\#Y_{n,x}$ follows a (non-degenerated) Poisson distribution, then $c_n = 1$. The case of overdispersion

(underdispersion), that is, $\nu_n > \beta_n$ ($\nu_n < \beta_n$) corresponds to $c_n > 1$ ($c_n < 1$). Denote by δ_0 the Dirac delta function defined on $\mathbb{R}^d$. Recall that for any integrable real function h defined on $\mathbb{R}^d$, $h * \delta_0 = \delta_0 * h = h$, and for any $a \in \mathbb{R}$, $a\delta_0 * a\delta_0 = a^2\delta_0$, where we understand $0\delta_0$ as 0. Finally, let $\overset{n}{\underset{i=1}{*}} h_i = h_1 * \cdots * h_n$ where each h_i is either of the form $a_i\delta_0$, with a_i a real constant, or it is an integrable real function defined on $\mathbb{R}^d$.

Theorem 3.3 . Suppose g_{G_0} and $g_{G_{Z_n}}$ are of the form (B.9), that is, $g_{G_0} - 1 = af_0 * \tilde{f}_0$ and $g_{Z_n} - 1 = b_n f_{Z_n} * \tilde{f}_{Z_n}$ for $n = 1, 2, \ldots$. Then, for all $u \in \mathbb{R}^d$ and $n = 1, 2, \ldots$,

$$g_{G_n}(u) - 1 = \left(\frac{\rho_{G_0}}{\rho_{G_n}}\right)^2 af_0 * \tilde{f}_0 * \overset{n}{\underset{i=1}{*}} \Big\{ \big(\beta_i p_i f_i + q_i\delta_0\big) * \big(\beta_i p_i \tilde{f}_i + q_i\delta_0\big) \Big\}(u) \tag{B.11}$$

$$+ \sum_{i=1}^{n} \frac{\rho_{G_{i-1}}}{\rho_{G_n}^2} \Big\{ c_i \big(\beta_i p_i\big)^2 f_i * \tilde{f}_i + \beta_i p_i q_i \big(f_i + \tilde{f}_i\big) \Big\} * \overset{n}{\underset{j=i+1}{*}} \Big\{ \big(\beta_j p_j f_j + q_j\delta_0\big) * \big(\beta_j p_j \tilde{f}_j + q_j\delta_0\big) \Big\}(u) \tag{B.12}$$

$$+ \sum_{i=1}^{n-1} \left(\frac{\rho_{Z_i}}{\rho_{G_n}}\right)^2 b_i f_{Z_i} * \tilde{f}_{Z_i} * \overset{n}{\underset{j=i+1}{*}} \Big\{ \big(\beta_j p_j f_j + q_j\delta_0\big) * \big(\beta_j p_j \tilde{f}_j + q_j\delta_0\big) \Big\}(u) \tag{B.13}$$

$$+ \left(\frac{\rho_{Z_n}}{\rho_{G_n}}\right)^2 b_n f_{Z_n} * \tilde{f}_{Z_n}(u) \tag{B.14}$$

where the $(n-i)$-th fold convolution in (B.12) is interpreted as 1 if $i = n$ and the sum in (B.13) is interpreted as 0 if $n = 1$.

The terms in (B.11)–(B.14) have the following interpretation when $u \neq 0$: The right side of (B.11) corresponds to pairs of n-th generation points with different 0-th generation ancestors; the i-th term in (B.12) corresponds to pairs of n-th generation points when they have a common $(i-1)$-th generation ancestor initiated by Z_{i-1} if $i > 1$ or by G_0 if $i = 1$; the i-th term in (B.13) corresponds to pairs of n-th generation points with different i-th generation ancestors initiated by Z_i; and (B.14) corresponds to point pairs in Z_n.

Later in Section 4.1, our main interest is in the behaviour of g_{G_n} as $n \to \infty$ when we have the same reproduction system, but for the moment, it is worth noticing the flexibility of our model for G_1 and the effect of the choice of its centre process G_0: For simplicity, suppose there is no noise and no retention of points from one generation to the next (i.e., $\rho_{Z_n} = q_n = 0$ for

$n = 1, 2, \ldots$), and G_0 is stationary and either a Poisson or an weighted determinantal or permanental point process with a Gaussian kernel. Specifically, suppose $d = 2$, G_0 has intensity $\rho_{G_0} = 100$, and using a notation as in Appendix A, the Gaussian kernel has an auto-correlation function of the form $R(x) = \exp(-\|x/\tau\|^2)$, where the value of τ depends on the type of process: For the α-weighted determinantal point process, we consider the most repulsive case, that is, a determinantal point process ($\alpha = 1$) and $\tau = 1/\sqrt{\rho_{G_0}\pi}$ is largest possible to ensure existence of the process Lavancier et al. (2015); for the α-weighted permanental point process, $\alpha = 1/2$ (the most attractive case when it is also a Cox process, see Appendix A) and $\tau = 0.1$ is an arbitrary value (any positive number can be used). Note that $R^2 = (\sqrt{\pi}\tau)^2 f_0 * \tilde{f}_0$ where $f_0 \sim N_2(\tau^2/8)$, which by (B.30) and (B.31) mean that (B.9) is satisfied with $a = 2(\sqrt{\pi}\tau)^2$ and $a = -(\sqrt{\pi}\tau)^2$ for the weighted permanental and determinantal point processes, respectively, and $a = 0$ in case of the Poisson process. Moreover, let the number of points in a cluster be Poisson distributed with mean $\beta_1 = 10$, $p_1 = 1$, and $f_1 \sim N_2(\sigma^2)$, with $\sigma = 0.01$. Then, by Theorem 3.3,

$$g_{G_1}(u) - 1 = \frac{a}{2\pi(2\sigma^2 + \tau^2/4)} \exp\left\{-\frac{\|u\|^2}{2(2\sigma^2 + \tau^2/4)}\right\} + \frac{1}{\rho_{G_0} 4\pi\sigma^2} \exp\left(-\frac{\|u\|^2}{4\sigma^2}\right).$$

In Figure B.1 we present the isotropic PCF $g_{G_1,o}(r) = g_{G_1}(u)$ as a function of the inter-point distance $r = \|u\|$ in case of each of the three models of G_0, where using an obvious notation, $g^{\text{det}}_{G_1,o} < g^{\text{Pois}}_{G_1,o} < g^{\text{wper}}_{G_1,o}$. Most notable is the fact that $g^{\text{det}}_{G_1,o}(r)$ exhibits repulsion at midrange distances r. For $g^{\text{wper}}_{G_1,o}$, we see a high degree of clustering, which is persistent for large values of r; this will of course be even more pronounced if we increase the value of τ; whilst decreasing σ will increase the peak at small values of r. Figure B.2 shows simulations of G_1 in each of the three cases of the model of G_0. As expected, we clearly see a higher degree of repulsion when G_0 is a determinantal point process (the left most plot) and a higher degree of clustering when G_0 is a weighted permanental point process (the right most plot). In particular, the clusters are more distinguishable when G_0 is a determinantal point process, and this will be even more pronounced if decreasing σ because the spread of clusters then decrease.

Proof of Theorem 3.3

Shimatani in Shimatani (2010) verified Theorem 3.3 in the special case where $q_1 = q_2 = \cdots = 0$, $b_1 = b_2 = \cdots = 0$ (as is the case if $Z_1, Z_2, \ldots$ are stationary Poisson processes), and $c_1 = c_2 = \cdots > 0$, in which case the terms

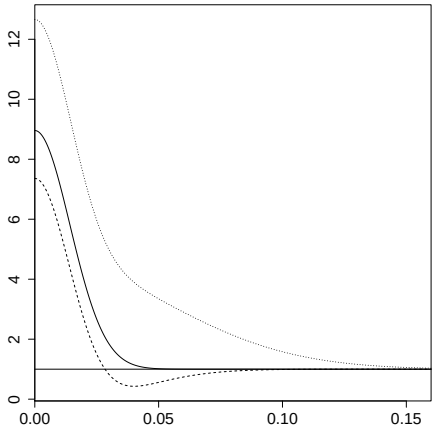

Fig. B.1: The PCFs of G_1 when G_0 is a determinantal, Poisson, or weighted permanental point process (dashed, solid, and dotted respectively), with parameters and Gaussian offspring PDF as specified in the text. The solid horizontal line is the PCF for a Poisson process.

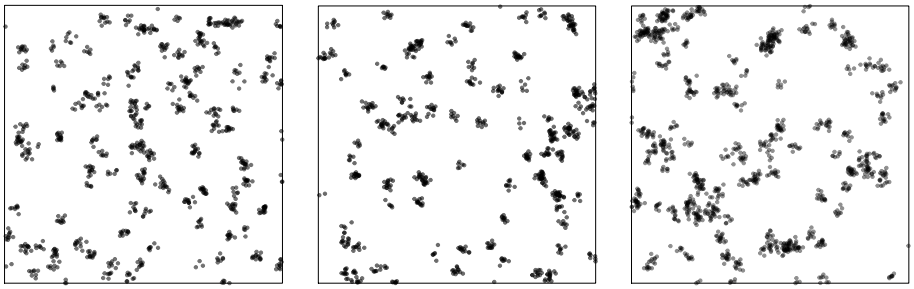

Fig. B.2: Simulations of G_1 restricted to a unit square when G_0 is a determinantal (left panel), Poisson (middle panel), or weighted permanental (right panel) point process, see Figure B.1 and the text.

in (B.13)–(B.14) are zero. If $c_1 = c_2 = \cdots = 0$, then (B.12) is zero and by (B.10), with probability one, $\#Y_{n,x} \in \{0, 1\}$ for all $x \in G_{n-1}$ and $n = 1, 2, \ldots$. Consequently, the proof of Theorem 3.3 is trivial if $c_1 = c_2 = \cdots = 0$ and both G_0 and $Z_1, Z_2, \ldots$ are stationary Poisson processes, because then $a = 0$, $b_1 = b_2 = \cdots = 0$, $G_1, G_2, \ldots$ are stationary Poisson processes, and the class of stationary Poisson processes is closed under IID random shifts of the points, thinning, and superposition. The general proof of Theorem 3.3 follows by induction from the following lemma together with Lemma 3.2.

Lemma 3.4 . If $\rho_{G_{n-1}} > 0$, $\rho_{G_n} > 0$, and $g_{G_{n-1}}$ and g_{Z_n} exist, then g_{G_n} exists

and is given by

$$\begin{aligned}
&g_{G_n}(u-v) - 1 \\
&\quad = \left(\frac{\rho_{G_{n-1}}\beta_n p_n}{\rho_{G_n}}\right)^2 \left[\iint \{g_{G_{n-1}}(x_1 - x_2) - 1\} f_n(u - x_1) f_n(v - x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \right. \\
&\qquad\qquad \left. + \frac{c_n}{\rho_{G_{n-1}}} f_n * \tilde{f}_n(u-v)\right] \\
&\quad + \left(\frac{\rho_{G_{n-1}} q_n}{\rho_{G_n}}\right)^2 \{g_{G_{n-1}}(u-v) - 1\} + \left(\frac{\rho_{Z_n}}{\rho_{G_n}}\right)^2 \{g_{Z_n}(u-v) - 1\} \\
&\quad + \frac{\rho_{G_{n-1}}^2 \beta_n p_n q_n}{\rho_{G_n}^2} \left[\int \{g_{G_{n-1}}(v-x) - 1\} \{f_n(u-x) + \tilde{f}_n(u-x)\}\,\mathrm{d}x \right. \\
&\qquad\qquad \left. + \frac{1}{\rho_{G_{n-1}}} \{f_n(u-v) + \tilde{f}_n(u-v)\}\right]
\end{aligned}$$

for any $u, v \in \mathbb{R}^d$.

Proof. Note that Y_n is stationary with intensity

$$\rho_{Y_n} = \rho_{G_{n-1}} \beta_n. \tag{B.15}$$

It follows straightforwardly from (B.1), (B.7), and Fubini's theorem that its PCF is given by

$$\begin{aligned}
\rho_{Y_n}^2 g_{Y_n}(u-v) = \rho_{G_{n-1}}^2 \beta_n^2 \iint g_{G_{n-1}}(x_1 - x_2) f_n(u - x_1) f_n(v - x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
+ \rho_{G_{n-1}} c_n \beta_n^2 f_n * \tilde{f}_n(u-v)
\end{aligned} \tag{B.16}$$

for any $u, v \in \mathbb{R}^d$, where the two terms on the right hand side correspond to pairs of points from Y_n belonging to different clusters and the same cluster, respectively. Hence by (B.2) and (B.15), W_n is stationary with intensity

$$\rho_{W_n} = p_n \rho_{Y_n} = \rho_{G_{n-1}} \beta_n p_n \tag{B.17}$$

and PCF

$$\begin{aligned}
g_{W_n}(u-v) &= g_{Y_n}(u-v) \\
&= \iint g_{G_{n-1}}(x_1 - x_2) f_n(u - x_1) f_n(v - x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&\quad + \frac{c_n}{\rho_{G_{n-1}}} f_n * \tilde{f}_n(u-v)
\end{aligned} \tag{B.18}$$

where the first identify follows from the fact that PCFs are invariant under independent thinning, and where (B.16) is used to obtain the second identity.

Also, for disjoint Borel sets $A_1, A_2 \subseteq \mathbb{R}^d$, it follows from items (a)–(c) that

$$\begin{aligned}&\mathrm{E}\{\#(W_n \cap A_1)\#(G_{n-1}^{\mathrm{thin}} \cap A_2)\}\\ &\quad = \rho_{W_n}\rho_{G_{n-1}^{\mathrm{thin}}} \int_{A_1}\int_{A_2} \left\{ \frac{1}{\rho_{G_{n-1}}} f_n(x_1 - x_2) + g_{G_{n-1}} * \tilde{f}_n(x_1 - x_2) \right\} \mathrm{d}x_1 \mathrm{d}x_2.\end{aligned}$$

Furthermore, by (B.3), (B.7), and Fubini's theorem it is readily seen that G_n has PCF given by

$$\begin{aligned}\rho_{G_n}^2 g_{G_n}(x) = \rho_{W_n}^2 g_{W_n}(x) &+ \left(\rho_{G_{n-1}}^{\mathrm{thin}}\right)^2 g_{G_{n-1}^{\mathrm{thin}}}(x) + \rho_{Z_n}^2 g_{Z_n}(x)\\ &+ 2\rho_{W_n}\rho_{Z_n} + 2\rho_{Z_n}\rho_{G_{n-1}}^{\mathrm{thin}}\\ &+ 2\rho_{W_n}\rho_{G_{n-1}^{\mathrm{thin}}} \left\{ g_{G_{n-1}} * f_n(x) + \frac{1}{\rho_{G_{n-1}}} f_n(x) \right\}\end{aligned}$$

where the six terms on the right hand side correspond to pairs of points from W_n, Z_n, G_{n-1}^{thin}, W_n and Z_n, Z_n and G_{n-1}^{thin}, and W_n and G_{n-1}^{thin}, respectively, where the latter three cases can be ordered in two ways. Combining all this with the first identity in (B.6) and (B.17), we easily obtain

$$\begin{aligned}&g_{G_n}(u - v) - 1\\ &\quad = \left(\frac{\rho_{G_{n-1}}\beta_n p_n}{\rho_{G_n}}\right)^2 \{g_{W_n}(u - v) - 1\} + \left(\frac{\rho_{Z_n}}{\rho_{G_n}}\right)^2 \{g_{Z_n}(u - v) - 1\}\\ &\qquad + \left(\frac{\rho_{G_{n-1}} q_n}{\rho_{G_n}}\right)^2 \{g_{G_{n-1}}(u - v) - 1\}\\ &\qquad + \rho_{W_n}\rho_{G_{n-1}^{\mathrm{thin}}} \Bigg[\int (g_{G_{n-1}}(v - x) - 1) \left\{ f_n(u - x) + \tilde{f}_n(u - x) \right\} \mathrm{d}x\\ &\qquad\qquad + \frac{1}{\rho_{G_{n-1}}} \left\{ f_n(u - v) + \tilde{f}_n(u - v) \right\} \Bigg].\end{aligned}$$

This combined with (B.18) imply the result in Lemma 3.4. □

Extension

More generally than in Section 3.2 we may consider the case where the PCF of the initial generation G_0 and the noise Z_n are affine expressions:

$$g_{G_0} - 1 = a_0 + a_1 f_{0,1} * \tilde{f}_{0,1} + \cdots + a_k f_{0,k} * \tilde{f}_{0,k} \tag{B.19}$$

and

$$g_{Z_n} - 1 = b_{n,0} + b_{n,1} f_{Z_n,1} * \tilde{f}_{Z_n,1} + \cdots + b_{n,l} f_{Z_n,l} * \tilde{f}_{Z_n,l}, \qquad n = 1, 2, \ldots, \tag{B.20}$$

for real constants $a_0, \dots, a_k, b_{n,1}, \dots, b_{n,l}$ and PDFs $f_{0,1}, \dots, f_{0,k}, f_{Z_n,1}, \dots, f_{Z_n,l}$. For instance, the superposition of k independent Poisson, weighted permanental, or weigthed determinantal point processes has a PCF of the form (B.19) or (B.20). Also Theorem 3.3 provides examples of PCFs of the form (B.19) or (B.20). Assuming (B.19) and (B.20), Theorem 3.3 is immediately generalised by replacing $af_0 * \tilde{f}_0$ in (B.11) by (B.19), $b_n f_{Z_n} * \tilde{f}_{Z_n}$ in (B.14) by (B.20), and similarly for $b_i f_{Z_i} * \tilde{f}_{Z_i}$ in (B.13).

4 Same reproduction system

Throughout this section we assume the same reproduction system over generations, that is, in items (a)–(d), $\beta_n = \beta$, $\nu_n = \nu$, $f_n = f$, $p_n = p$, $q_n = q$ do not depend on n, $Z_1, Z_2, \dots$ are IID stationary point processes, so $\rho_{Z_n} = \rho_Z$ for $n = 1, 2, \dots$, and $\rho_{G_0} = \rho_{G_1} = \cdots = \rho_G > 0$. Note that the noise process Z_n and the initial generation process G_0 need not be Poisson processes and the offspring densities need not be Gaussian as in Shimatani's paper Shimatani (2010). By (B.6), we have either

$$\beta p + q = 1 \quad \text{and} \quad \rho_Z = 0, \tag{B.21}$$

or

$$\beta p + q < 1 \quad \text{and} \quad \rho_Z > 0. \tag{B.22}$$

In case of (B.22),

$$\rho_G = \rho_Z / (1 - \beta p - q). \tag{B.23}$$

4.1 Limiting pair correlation function

Under the assumptions above and in Theorem 3.3, setting $0^0 = 1$, the PCF simplifies after a straightforward calculation to

$$\begin{aligned} g_{G_n}(u) - 1 = {} & af_0 * \tilde{f}_0 * \sum_{k_1=0}^{n} \sum_{k_2=0}^{n} \binom{n}{k_1} \binom{n}{k_2} q^{2n-k_1-k_2} (\beta p)^{k_1+k_2} f^{*k_1} * \tilde{f}^{*k_2}(u) \\ & + \left\{ \frac{c(\beta p)^2 f * \tilde{f} + \beta p q (f + \tilde{f})}{\rho_G} + \left(\frac{\rho_Z}{\rho_G} \right)^2 b f_Z * \tilde{f}_Z \right\} \\ & * \sum_{i=0}^{n-1} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \binom{i}{k_1} \binom{i}{k_2} q^{2i-k_1-k_2} (\beta p)^{k_1+k_2} f^{*k_1} * \tilde{f}^{*k_2}(u), \end{aligned} \tag{B.24}$$

for $n = 1, 2, \dots$, where

$$c = (\nu + \beta^2 - \beta)/\beta^2 \quad \text{if } \beta > 0, \qquad c = 0 \quad \text{if } \beta = 0,$$

f^{*n} is the n-th convolution power of f if $n > 0$, and $f^{*0} * \tilde{f}^{*0} = \delta_0$. For instance, consider the case $f \sim N_d(\sigma^2)$ and $f_Z \sim N_d(\kappa^2)$, and suppose $d \geq 3$ in case of (B.21). Then the binomial formula combined with either (B.21) or (B.22) imply that the first double sum in (B.24) tends to 0 as $n \to \infty$, and hence

$$\begin{aligned} g_G(u) - 1 &:= \lim_{n\to\infty} g_{G_n}(u) - 1 \\ &= \frac{c}{\rho_G} \sum_{i=0}^{\infty} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \binom{i}{k_1}\binom{i}{k_2} \frac{q^{2i-k_1-k_2}(\beta p)^{2+k_1+k_2}}{\{2\pi(2+k_1+k_2)\sigma^2\}^{d/2}} \\ &\qquad \cdot \exp\left\{-\frac{\|u\|^2}{2(2+k_1+k_2)\sigma^2}\right\} \\ &\quad + \frac{2}{\rho_G} \sum_{i=0}^{\infty} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \binom{i}{k_1}\binom{i}{k_2} \frac{q^{2i-k_1-k_2+1}(\beta p)^{1+k_1+k_2}}{\{2\pi(1+k_1+k_2)\sigma^2\}^{d/2}} \\ &\qquad \cdot \exp\left\{-\frac{\|u\|^2}{2(1+k_1+k_2)\sigma^2}\right\} \\ &\quad + \left(\frac{\rho_Z}{\rho_G}\right)^2 \sum_{i=0}^{\infty} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \binom{i}{k_1}\binom{i}{k_2} \frac{q^{2i-k_1-k_2}(\beta p)^{k_1+k_2}}{\left[2\pi\left\{(k_1+k_2)\sigma^2 + 2\kappa^2\right\}\right]^{d/2}} \\ &\qquad \cdot \exp\left[-\frac{\|u\|^2}{2\left\{(k_1+k_2)\sigma^2 + 2\kappa^2\right\}}\right] \end{aligned} \tag{B.25}$$

is finite. Shimatani in Shimatani (2010) only showed that this is finite under the assumption $d = 2$, $b = q = 0$, and $c > 0$. Then Shimatani noticed that $\beta p = 1$ and $\rho_Z = 0$ (which is (B.21) for $q = 0$) imply divergence of g_{G_n} as $n \to \infty$, whilst $\beta p < 1$ and $\rho_Z > 0$ (which is (B.22) for $q = 0$) imply convergence. Further, in the case of convergence and when $\beta p \approx 1$, he discussed an approximation of $g_G(u)$ that depends on whether $\|u\|$ is close to 0 or not.

In general (i.e., without making the assumption of normal distributions and so on), if we assume $g_{G_n} - 1$ has a finite limit and (B.22) is satisfied, then

$$\begin{aligned} g_G(u) - 1 = &\left\{\frac{c(\beta p)^2 f * \tilde{f} + \beta p q(f + \tilde{f})}{\rho_G} + \left(\frac{\rho_Z}{\rho_G}\right)^2 b f_Z * \tilde{f}_Z\right\} \\ &* \sum_{i=0}^{\infty} \sum_{k_1=0}^{i} \sum_{k_2=0}^{i} \binom{i}{k_1}\binom{i}{k_2} q^{2i-k_1-k_2}(\beta p)^{k_1+k_2} f^{*k_1} * \tilde{f}^{*k_2}(u) \end{aligned} \tag{B.26}$$

which does not depend on a or f_0. Here, as $\beta p + q \uparrow 1$, ρ_Z/ρ_G goes to 0, meaning that the less noise we consider, the less it matters which type of PCF for the noise process Z_n we choose. On the other hand, as $\beta p + q \downarrow 0$, $g_G - 1$ tends to $b f_Z * \tilde{f}_Z$, which simply is the PCF of Z_n.

Considering the situation at the end of Section 3.2, assume that $d = 2$, $q = 0$, $f \sim N_d(\sigma^2)$, and $g_{Z_n} - 1 = b f_Z * \tilde{f}_Z$ (corresponding to (B.9)) with

$f_Z \sim N_d(\kappa^2/8)$ and $b = 0$, $b = -(\sqrt{\pi}\kappa)^2$, and $b = 2(\sqrt{\pi}\kappa)^2$ for the Poisson, determinantal, and weighted permanental point process, respectively. Then $g_G(u)$ is given by (B.25), where $d = 2$ and κ^2 is replaced by $\kappa^2/8$. Also assume that $p = 1$, $\sigma = 0.1$, $\rho_G = 100$, and the number of points in a cluster is Poisson distributed (implying $c = 1$) with mean $\beta = 0.8$, so $\rho_Z = 20$. Finally, assume $\kappa = 0.1$ in case of weighted permanental noise and $\kappa = 1/\sqrt{\rho_Z \pi}$ in case of determinantal noise (the most repulsive Gaussian determinantal point process). Shimatani in Shimatani (2010) discussed the case where $\beta p = 0.99$ – a plot (omitted here) shows that the limiting PCFs corresponding to the three models of noise processes are then effectively equal. By lowering βp, the reproduction system is diminished, and hence depending on the model type, a higher degree of regularity or clustering is obtained. This will also increase the rate of convergence because the number of generations initialized by a single point will be fewer. Note that in Figure B.3 the convergence is already rapid as g_{G_8} and $g_{G_{16}}$ are practically indistinguishable. Figure B.3 further shows that it is only for small or moderate inter-point distances that the three limiting PCFs clearly differ.

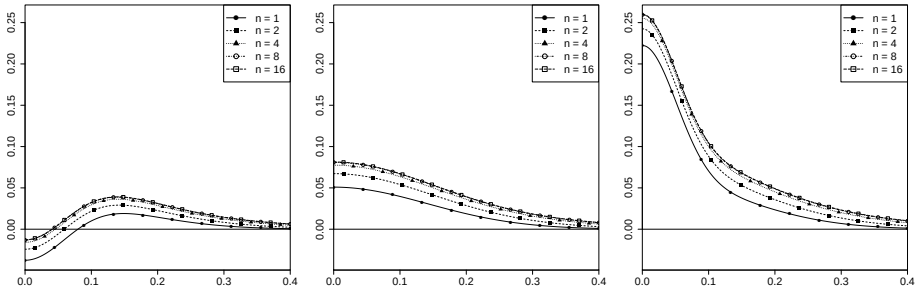

Fig. B.3: The reduced PCFs $g_{G_n} - 1$ when the noise processes are either determinantal, Poisson or weighted permanental point processes (left to right), with parameters and Gaussian offspring PDF as specified in the text. The solid horizontal line is the PCF - 1 for a Poisson process.

4.2 Second main result

Although Shimatani in Shimatani (2010) showed convergence of g_{G_n} in the special case considered above, he did not clarify whether the Markov chain $G_0, G_1, \ldots$ converges in distribution to a limit so that this limiting distribution (also called the equilibrium, invariant, or stationary distribution) has a PCF given by (B.26). In order to show that $G_0, G_1, \ldots$ is indeed converging to a limiting distribution under more general conditions, and to specify what this is, we construct in accordance with items (a)–(d) a Markov chain $\ldots, G^{\mathrm{st}}_{-1}, G^{\mathrm{st}}_0, G^{\mathrm{st}}_1, \ldots$ with times given by all integers n and so that this chain is time-stationary (its distribution is invariant under discrete time shifts), as follows. First, we generate noise processes as in item (d): Let $\ldots, Z_{-1}, Z_0, Z_1, \ldots$

be independent stationary Poisson processes on $\mathbb{R}^d$ with intensity ρ_Z. Second, for any integer n and point $x \in Z_n$, we consider the family of all generations initiated by the ancestor x, that is, the family

$$F_{n,x} = \bigcup_{m=1}^{\infty} W_{n,x}^{(m)}$$

where

$$W_{n,x}^{(1)} = \begin{cases} W_{n,x} \cup \{x\} & \text{if } Q_{1,x} = 1, \\ W_{n,x} & \text{if } Q_{1,x} = 0, \end{cases}$$

is defined by the reproduction mechanism of independent clustering, independent thinning, and independent retention given in items (a)–(c) (with $\beta_n = \beta$ and $\nu_n = \nu$), $W_{n,x}^{(2)}$ is the offspring and retained points generated by the points in $W_{n,x}^{(1)}$ (using the same reproduction mechanism as before), and so on. In other words, $W_{n,x}^{(m)}$ is the set of $(m+n)$-th generation points with common n-th generation ancestor $x \in Z_n$. Moreover, we assume that conditional on $\ldots, Z_{-1}, Z_0, Z_1, \ldots$, the families $F_{n,x}$ for all integers n and $x \in Z_n$ are independent (and hence IID). Finally, for all integers n, we let

$$G_n^{\mathrm{st}} = W_n^{\mathrm{st}} \cup Z_n \qquad \text{with } W_n^{\mathrm{st}} = \bigcup_{m=1}^{\infty} \bigcup_{x \in Z_{n-m}} W_{n-m,x}^{(m)}. \tag{B.27}$$

It will be evident from the next theorem that any G_n^{st} has intensity ρ_G given by (B.23) and PCF g_G given by (B.26) (provided $g_G(u-v)$ is a locally integrable function of $(u,v) \in \mathbb{R}^d \times \mathbb{R}^d$); a formal proof is given in Appendix B. The proof of Theorem 4.2 is based on a coupling construction between $G_1, G_2, \ldots$ and $G_1^{\mathrm{st}}, G_2^{\mathrm{st}}, \ldots$ together with the following result.

Lemma 4.1 . Suppose $\beta_n = \beta$, $\nu_n = \nu$, $f_n = f$, $p_n = p$, $q_n = q$, and $\rho_{Z_n} = \rho_Z$ do not depend on $n \geq 1$, where $\beta p + q < 1$ and $\rho_Z > 0$. Let $K \subset \mathbb{R}^d$ be a compact set and let

$$T_{0,K}^{\mathrm{st}} = \sup\bigl\{m \in \{1,2,\ldots\} : W_{0,x}^{(m)} \cap K \neq \emptyset \text{ for some } x \in G_0^{\mathrm{st}}\bigr\} \tag{B.28}$$

be the last time a point in K is a member of a family initiated by some point in the 0-th generation G_0^{st}. Then

$$\mathrm{E}\left(T_{0,K}^{\mathrm{st}}\right) \leq |K| \rho_G \frac{\beta p + q}{1 - \beta p - q}$$

is finite, and so $T_{0,K}^{\mathrm{st}} < \infty$ almost surely.

Proof. Let $K \subset \mathbb{R}^d$ be compact and define

$$N = \sum_{x \in G_0^{\mathrm{st}}} \#(F_{0,x} \cap K).$$

By the law of total expectation, conditioning on G_0^{st} and using Campbell's theorem, we obtain

$$\begin{aligned}\mathrm{E}(N) &= \rho_G \int \sum_{m=1}^{\infty} \int_K \bigl(\beta p f + q\delta_0\bigr)^{*m}(y-x)\,\mathrm{d}y\,\mathrm{d}x \\ &= \rho_G \int \sum_{m=1}^{\infty} \int_K \sum_{k=0}^{m} \binom{m}{k} q^{m-k}(\beta p)^k f^{*k}(y-x)\,\mathrm{d}y\,\mathrm{d}x \\ &= |K|\rho_G \frac{\beta p + q}{1-\beta p - q} \end{aligned} \tag{B.29}$$

using Fubini's theorem in the last identity. Further, the families initiated by the points in G_0^{st} are almost surely pairwise disjoint, so N is almost surely the number of points in K belonging to some family initiated by a point $x \in G_0^{\mathrm{st}}$. Consequently, $\mathrm{P}(T_{0,K}^{\mathrm{st}} \le N) = 1$, whereby the lemma follows. □

We are now ready to state our second main result.

Theorem 4.2 . Suppose $\ldots, Z_{-1}, Z_0, Z_1 \ldots$ are IID stationary point processes and $\beta_n = \beta$, $\nu_n = \nu$, $f_n = f$, $p_n = p$, $q_n = q$, and $\rho_{Z_n} = \rho_Z$ do not depend on $n \ge 1$, where $\beta p + q < 1$ and $\rho_Z > 0$. Then $\ldots, G_{-1}^{\mathrm{st}}, G_0^{\mathrm{st}}, G_1^{\mathrm{st}}, \ldots$ is a time-stationary Markov chain constructed in accordance to items (a)–(d). Let Π be the distribution of any G_n^{st} and let $\mathcal{N}$ be the space of all locally finite subsets of $\mathbb{R}^d$. Then there exists a (measurable) subset $\Omega \subseteq \mathcal{N}$ so that $\Pi(\Omega) = 1$ and for any compact set $K \subset \mathbb{R}^d$ and all $\omega \in \Omega$, conditional on $G_0 = \omega$, there is a coupling between $G_1, G_2, \ldots$ and $\ldots, G_{-1}^{\mathrm{st}}, G_0^{\mathrm{st}}, G_1^{\mathrm{st}}, \ldots$, and there exists a random time $T_K(\omega) \in \{0, 1, \ldots\}$ so that $G_n \cap K = G_n^{\mathrm{st}} \cap K$ for all integers $n > T_K(\omega)$. In particular, for any $\omega \in \Omega$ and conditional on $G_0 = \omega$, G_n converges in distribution to Π as $n \to \infty$, and so Π is the unique invariant distribution of the chain $G_0, G_1, \ldots$.

Proof. Obviously, $\ldots, G_{-1}^{\mathrm{st}}, G_0^{\mathrm{st}}, G_1^{\mathrm{st}}, \ldots$ is a time-stationary Markov chain constructed in accordance to items (a)–(d). To verify the remaining part of the theorem, we may assume that G_0 and G_0^{st} are independent. Then, conditional on G_0, we have a coupling between $G_1, G_2, \ldots$ and $\ldots, G_{-1}^{\mathrm{st}}, G_0^{\mathrm{st}}, G_1^{\mathrm{st}}, \ldots$ because $G_1^{\mathrm{st}}, G_2^{\mathrm{st}}, \ldots$ and $G_1, G_2, \ldots$ are generated by the same noise processes $Z_1, Z_2, \ldots$, the same offspring processes $Y_{n,x}$ for all times $n = 1, 2, \ldots$ and all ancestors $x \in G_{n-1} \cap G_{n-1}^{\mathrm{st}}$, the same Bernoulli variables $B_{n,y}$ for all times $n = 1, 2, \ldots$ and all offspring $y \in Y_{n,x}$ with ancestor $x \in G_{n-1} \cap G_{n-1}^{\mathrm{st}}$, and the same Bernoulli variables $Q_{n,x}$ for all times $n = 1, 2, \ldots$ and all retained points $x \in G_{n-1} \cap G_{n-1}^{\mathrm{st}}$. Let $K \subset \mathbb{R}^d$ be compact. In accordance with (B.28), for $\omega \in \mathcal{N}$, let

$$T_K(\omega) = \sup\bigl\{m \in \{1, 2, \ldots\} : W_{0,x}^{(m)} \cap K \ne \emptyset \text{ for some } x \in \omega\bigr\}$$

be the last time a point in K is a member of a family initiated by some point in ω, and let $\Omega = \{\omega \in \mathcal{N} : T_K(\omega) < \infty\}$. By Lemma 4.1 and the coupling construction, $\Pi(\Omega) = 1$ and $G_n \cap K = G_n^{\text{st}} \cap K$ whenever $n > T_K(\omega)$, so for any $\omega \in \Omega$,

$$\lim_{n\to\infty} \mathrm{P}\left(G_n \cap K = \emptyset | G_0 = \omega\right) = \lim_{n\to\infty} \mathrm{P}\left(G_n^{\text{st}} \cap K = \emptyset, n > T_K(\omega)\right)$$

because G_0 is independent of $(G_0^{\text{st}}, T_K(\omega))$. Since the sequence of events $\{\omega : 1 > T_K(\omega)\} \subseteq \{\omega : 2 > T_K(\omega)\} \subseteq \ldots$ increases to Ω, we obtain

$$\lim_{n\to\infty} \mathrm{P}\left(G_n \cap K = \emptyset | G_0 = \omega\right) = \lim_{n\to\infty} \mathrm{P}\left(G_n^{\text{st}} \cap K = \emptyset\right) = \mathrm{P}\left(G_0^{\text{st}} \cap K = \emptyset\right).$$

Thus, recalling that the distribution of a random closed set $X \subseteq \mathbb{R}^d$ (e.g. a locally finite point process) is uniquely characterized by the void probabilities $\mathrm{P}(X \cap K = \emptyset)$ for all compact sets $K \subset \mathbb{R}^d$, we have verified that conditional on $G_0 = \omega$, the chain $G_1, G_2 \ldots$ converges in distribution towards Π. In turn, this implies uniqueness of the invariant distribution Π. □

In Theorem 4.2, under mild conditions, we can take $\Omega = \mathcal{N}$. For instance, this is easily seen to be the case if there exists $\varepsilon > 0$ so that $f(x) > 0$ whenever $\|x\| \leq \varepsilon$. In the special case $c = 0$, Π is just a stationary Poisson process, and so $\Omega = \mathcal{N}$. Moreover, the integral

$$\gamma := \int (g_G - 1)$$

is a rough measure of the amount of positive/negative association between the points in G_n^{st}. Note that comparing γ with the corresponding measure for another stationary point process makes only sense if the processes have equal intensities, see Lavancier et al. (2015). Under the assumptions in both Theorem 3.3 and 4.2, by (B.26),

$$\begin{aligned}\gamma &= \frac{c(\beta p)^2 + 2\beta pq}{\rho_G \{1 - (\beta p + q)^2\}} + \frac{b\rho_Z^2}{\rho_G^2 \{1 - (\beta p + q)^2\}} \\ &= \frac{1}{1 + \beta p + q} \left\{ \frac{c(\beta p)^2 + 2\beta pq}{\rho_Z} + b(1 - \beta p - q) \right\}\end{aligned}$$

which does not depend on f or f_Z. Furthermore, γ may take any positive value and some negative values depending on how we choose the values of the parameters. This means we may have an equilibrium distribution exhibiting any degree of clustering or some degree of regularity. In fact, γ can only be negative when b is negative, e.g when Z_n is a determinantal point process. In this case b has a lower bound, $b_{\min}$, that ensures the existence of the determinantal point process Lavancier et al. (2015) and consequently, $\gamma \geq b_{\min}$. The case $\gamma = b_{\min}$ happens exactly when $\beta p + q = 0$ (i.e., when offspring

are never produced or no points are retained after the thinning procedures in items (b) and (c)) and thus $G_n = Z_n$ is a determinantal point process.

For approximate simulation of G_0^{st} under each of the three models of the noise processes, we use the algorithm described in Appendix C. Simulation was initially done with parameters and set-up corresponding to that of Figure B.3. However, the resulting point patterns were not distinguishable from a stationary Poisson process when comparing empirical estimates of the PCF, L-function, or J-function of the simulations to 95% global rank envelopes under each model (for definition of L- and J-functions, see e.g. Møller and Waagepetersen (2004), and for the envelopes, see Myllymäki et al. (2017)). Therefore, in order to better distinguish the three models, we consider two cases as follows.

Case 1:

This case is based on minimizing γ under determinantal noise and on maximizing γ under weighted permanental noise. Let $d = 2$, $f \sim N_d(\sigma^2)$, with $\sigma = 0.1$, $f_Z \sim N_d(\kappa^2/8)$, $\rho_G = 100$, $p = 1$, $\beta = 0.3$, $q = 0$, and consequently $\rho_Z = 70$.

- In case of determinantal noise: Let $\kappa = 1/\sqrt{\rho_Z \pi}$ (the most repulsive Gaussian determinantal point process) and the number of points in a cluster be Bernoulli distributed with parameter β, implying $c = 0$ (each point has at most one offspring). Then $\gamma \approx -5.38 \times 10^{-3}$.
- In case of Poisson noise: Let the number of points in a cluster be Poisson distributed with intensity β, implying $c = 1$. Then $\gamma \approx 9.89 \times 10^{-4}$.
- In case of weighted permanental noise: Let $\kappa = 1$ and the number of points in a cluster be negative binomially distributed with probability of success equal to 0.12 and dispersion parameter equal to 0.11, implying $c = 10$. Then $\gamma \approx 3.39$.

Case 2:

This case is such that the clusters are more separated. Let $d = 2$, $f \sim N_d(\sigma^2)$, with $\sigma = 0.01$, $f_Z \sim N_d(\kappa^2/8)$, $\rho_G = 100$, $p = 1$, $\beta = 0.95$, $q = 0$, and consequently $\rho_Z = 5$. Also, let the number of points in a cluster be negative binomially distributed with probability of success equal to 0.208 and dispersion parameter equal to 0.25, implying $c = 5$.

- In case of determinantal noise: Let $\kappa = 1/\sqrt{\rho_Z \pi}$. Then $\gamma \approx 0.463$.
- In case of Poisson noise: $\gamma \approx 0.463$.
- In case of weighted permanental noise: Let $\kappa = 1$. Then $\gamma \approx 0.624$.

Figure B.4 shows simulations of G_0^{st} under each of the three models of the noise processes (left to right) in Case 1 and 2 (top and bottom). Based on these simulations, Figure B.5 shows empirical estimates of functional summary statistics based on the simulated point patterns from Figure B.4 along with 95% global rank envelopes based on 2499 simulations (as recommended in Myllymäki et al. (2017)) of a stationary Poisson process with the same intensity as used in Figure B.4. The first simulated point pattern of Case 1 looks slightly less clustered than the second, whilst the last looks more clustered. This is in accordance with the values of γ and the corresponding functional summary statistics in Figure B.5. Additionally, Figure B.5 reveals that the case of Poisson noise is not distinguishable from the stationary Poisson process, while the case of weighted permanental noise is more clustered. The case of determinantal noise is not distinguishable from the stationary Poisson process by the PCF or L-function, but is shown to be more regular by the J-function. In Case 2, the clusters of the point pattern simulated under determinantal noise looks more separated than the clusters of the point pattern simulated under Poisson noise. The clusters of the point pattern simulated under weighted permanental noise are clustered to such a degree that it gives the illusion of few highly separated clusters. All three models of Case 2 are as expected significantly different from the stationary Poisson process.

Acknowledgements. Supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 "Statistics for point processes in space and beyond", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation. We thank Ina Trolle Andersen, Yongtao Guan, Ute Hahn, Henrike Häbel, Eva B. Vedel Jensen, and Morten Nielsen for helpful comments.

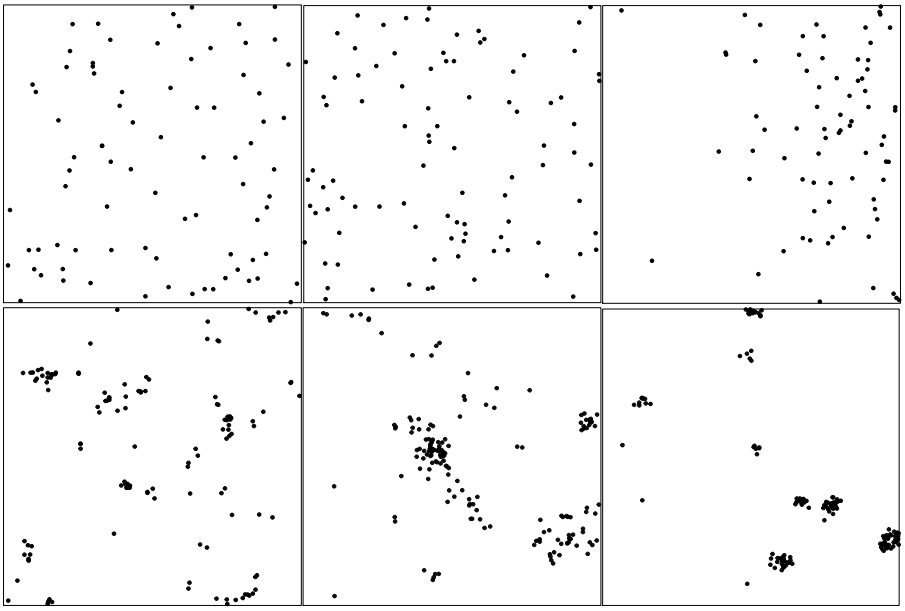

Fig. B.4: Simulations of G_0^{st} restricted to a unit square when the noise processes are either determinantal (left panel), Poisson (middle panel), or weighted permanental (right panel) point processes, with parameters as specified in the text. The rows corresponds to Case 1 and 2, respectively.

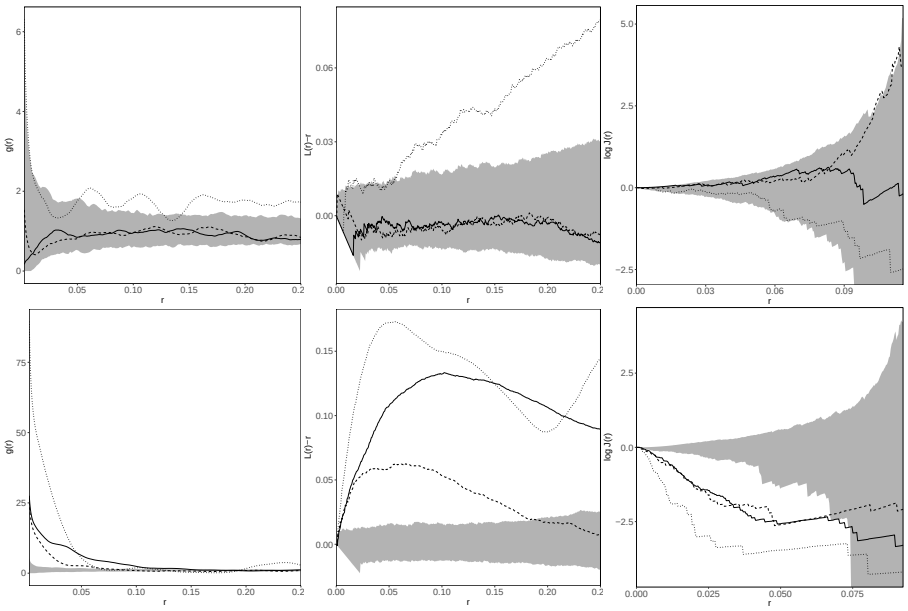

Fig. B.5: Empirical PCFs, L-functions, and J-functions (left to right) based on the simulations of G_0^{st} from Figure B.4 when the noise processes are either determinantal (dashed), Poisson (solid), or weighted permanental (dotted). The rows corresponds to Case 1 and 2, respectively. The grey regions are 95% global rank envelopes based on 2499 simulations of a stationary Poisson process with the same intensity as G_0^{st}.

A Weighted determinantal and permanental point processes

When defining stationary weighted determinantal/permanental point processes, the main ingredients are a symmetric function $C : \mathbb{R}^d \to \mathbb{R}$ and a real number α. Before giving the definitions of these point processes we recall the following.

For a real $n \times n$ matrix A with (i,j)-th entry $a_{i,j}$, the α-weighted permanent of A is defined by

$$\mathrm{per}_\alpha(A) = \sum_\sigma \alpha^{\#\sigma} a_{1,\sigma_1} \cdots a_{n,\sigma_n}$$

where σ denotes a permutation of $\{1,\dots,n\}$ and $\#\sigma$ is the number of its cycles. This is the usual permanent of A if $\alpha = 1$. Moreover, the α-weighted determinant of A is given by

$$\det_\alpha(A) = \mathrm{per}_{-\alpha}(-A).$$

This is the usual determinant of A if $\alpha = -1$. Often we just write $\mathrm{per}_\alpha A$ for $\mathrm{per}_\alpha(A)$, and $\det_\alpha A$ for $\det_\alpha(A)$.

For any $X_1,\dots,X_n \in \mathbb{R}^d$, the $n \times n$ matrix with (i,j)-th entry $C(X_i - X_j)$ is denoted by $[C](X_1,\dots,X_n)$. Thus

$$\mathrm{per}_\alpha[C](X_1,\dots,X_n) = \sum_\sigma \alpha^{\#\sigma} C(X_1 - X_{\sigma_1}) \cdots C(X_n - X_{\sigma_n}).$$

Note that the weighted permanent/determinant can be negative if the mapping $\mathbb{R}^d \times \mathbb{R}^d \ni (u,v) \to C(u-v)$ is not positive semi-definite. When this mapping is positive semi-definite, C is an auto-covariance function, with corresponding auto-correlation function $R(x) = C(x)/C(0)$ provided $C(0) > 0$.

A locally finite point process $X \subset \mathbb{R}^d$ has n-th order joint intensity $\rho_X^{(n)}$ for $n = 1,2,\dots$ if for any bounded and pairwise disjoint Borel sets $A_1,\dots,A_n \subset \mathbb{R}^d$,

$$\mathrm{E}\left[N(A_1)\cdots N(A_n)\right] = \int_{A_1} \cdots \int_{A_n} \rho_X^{(n)}(x_1,\dots,x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n < \infty.$$

Note that $\rho_X^{(n)}$ is unique except for a Lebesgue nullset in $\mathbb{R}^{dn}$ (we ignore nullsets in the following). Thus, if X is stationary, $\rho_X^{(1)}$ is constant and agrees with the intensity ρ_X, and $\rho_X > 0$ implies that $g_X(u-v) = \rho_X^{(2)}(u,v)/\rho_X^2$ is the PCF.

If for all $n = 1,2,\dots$, the n-th order joint intensity exists and is

$$\rho_X^{(n)}(X_1,\dots,X_n) = \mathrm{per}_\alpha[C](X_1,\dots,X_n)$$

we say that X is a stationary α-weighted permanental point process with kernel C and write $X \sim \mathrm{PPP}_\alpha(C)$. Conditions are needed to ensure the existence of $\mathrm{PPP}_\alpha(C)$, see Shirai and Takahashi (2003) and McCullagh and Møller (2006). To exclude the trivial case where X is empty we assume $\alpha C(0) > 0$. Note that C must be an auto-covariance function, $\alpha > 0$ since $\rho_X = \alpha C(0)$, and

$$g_X(x) - 1 = R(x)^2/\alpha. \tag{B.30}$$

This reflects that the process exhibits a positive association between its points. In fact, if C is an auto-covariance function and $k = 2\alpha$ is a positive integer, then $X \sim \mathrm{PPP}_\alpha(C)$ exists and it is a Cox process: Conditional on IID zero-mean stationary Gaussian processes $\Phi_1, \ldots, \Phi_k$ on $\mathbb{R}^d$ with auto-covariance function $C/2$, we can let X be a Poisson process with intensity function $\Lambda(x) = \Phi_1(x)^2 + \cdots + \Phi_k(x)^2$, $x \in \mathbb{R}^d$. In particular, if $\alpha = 1$, then X is the boson process introduced by Macchi (1975).

If for all $n = 1, 2, \ldots$, the n-th order joint intensity exists and is

$$\rho_X^{(n)}(G_1, \ldots, G_n) = \det_\alpha[C](G_1, \ldots, G_n)$$

we say that X is a stationary α-weighted determinantal point process with kernel C and write $X \sim \mathrm{DPP}_\alpha(C)$. To exclude the trivial case where X is empty we assume $\alpha C(0) > 0$. Again C needs to be an auto-covariance function, $\alpha > 0$ since $\rho_X = \alpha C(0)$, and

$$g_X(x) - 1 = -R(x)^2/\alpha. \tag{B.31}$$

If $\alpha = 1$, then X is the fermion process introduced by Macchi (1975) (it is usually called the determinantal point process). We have the following existence result: If C is continuous and square integrable, existence of $X \sim \mathrm{DPP}_1(C)$ is equivalent to the Fourier transform of C being bounded by 0 and 1 Lavancier et al. (2015). When α is a positive integer, $X \sim \mathrm{DPP}_\alpha(C)$ can be identified with the superposition $G_1 \cup \cdots \cup G_\alpha$ of independent processes $G_i \sim \mathrm{DPP}_\alpha(C/\alpha)$, $i = 1, \ldots, \alpha$. In general, the process is not well-defined if $0 < \alpha < 1$, cf. McCullagh and Møller (2006).

B The intensity and PCF of the invariant distribution

Let the situation be as in Theorem 4.2. Below we verify (B.23) and (B.26) holds for G_n^{st} provided $g_G(u - v)$ is a locally integrable function of $(u, v) \in \mathbb{R}^d \times \mathbb{R}^d$.

Note that the G_n^{st} are identically distributed and $G_0^{\mathrm{st}} = W_0^{\mathrm{st}} \cup Z_0$ where $W_0^{\mathrm{st}} = \bigcup_{m=1}^\infty \bigcup_{x \in Z_{-m}} W_{-m,x}^{(m)}$, cf. (B.27). Hence, for Borel sets $A \subseteq \mathbb{R}^d$ with

$|A| < \infty$, using similar arguments as in the derivation of (B.29), we obtain

$$\mathrm{E}\{\#(W_0^{\mathrm{st}} \cap A)\} = |A|\rho_Z \frac{\beta p + q}{1 - \beta p - q}, \tag{B.32}$$

so W_0^{st} has intensity

$$\rho_W = \rho_Z \frac{\beta p + q}{1 - \beta p - q} \tag{B.33}$$

whereby it follows that G_0^{st} has intensity ρ_G as given by (B.23).

Let $A_1, A_2 \subseteq \mathbb{R}^d$ be disjoint Borel sets with $|A_i| < \infty$, $i = 1, 2$. Using similar arguments as in the derivation of (B.29) (or (B.32)) and exploiting the fact that $Z_0, Z_{-1}, \ldots$ are IID point processes with a PCF of the form $g_Z = 1 + bf_Z * \tilde{f}_Z$ as well as the independence between Z_0 and W_0^{st}, we obtain

$$\begin{aligned}
&\mathrm{E}\{\#(G_0^{\mathrm{st}} \cap A_1)\#(G_0^{\mathrm{st}} \cap A_2)\} \\
&\quad = \rho_Z^2 |A_1||A_2| + \rho_Z^2 \int_{A_1} \int_{A_2} bf_Z * \tilde{f}_Z(x_1 - x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 + 2\rho_Z\rho_W|A_1||A_2| \qquad \text{(B.34)} \\
&\qquad + \sum_{m_1=1}^{\infty} \sum_{m_2=1:\, m_1 \neq m_2}^{\infty} \rho_Z^2 (\beta p + q)^{m_1+m_2} |A_1||A_2| \qquad \text{(B.35)} \\
&\qquad + \sum_{m=1}^{\infty} \rho_Z^2 (\beta p + q)^{2m} |A_1||A_2| \\
&\qquad + \sum_{m=1}^{\infty} \rho_Z^2 \int_{A_1} \int_{A_2} bf_Z * \tilde{f}_Z \\
&\qquad\quad * \sum_{k_1=0}^{m} \sum_{k_2=0}^{m} \binom{m}{k_1} \binom{m}{k_2} q^{2m-k_1-k_2} (\beta p)^{k_1+k_2} f^{*k_1} * \tilde{f}^{*k_2}(y_1 - y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 \qquad \text{(B.36)} \\
&\qquad + \sum_{m=1}^{\infty} \mathrm{E}\Big\{ \sum_{x \in Z_{-m}} \#(W_{-m,x}^{(m)} \cap A_1)\#(W_{-m,x}^{(m)} \cap A_2) \Big\}. \qquad \text{(B.37)}
\end{aligned}$$

Here,

- the first two terms of (B.34) corresponds to pairs of points from Z_0 with one point falling in A_1 and the other in A_2;
- the third term corresponds to pairs of points either from $Z_0 \cap A_1$ and $W_0^{\mathrm{st}} \cap A_2$ or from $Z_0 \cap A_2$ and $W_0^{\mathrm{st}} \cap A_1$;
- the term in (B.35) corresponds to pairs of points, with one point falling in A_1 and the other in A_2 of two families initiated by ancestors from different generations;

- the two terms in (B.36) corresponds to pairs of points, with one point falling in A_1 and the other in A_2 from two different families initiated by ancestors from the same generation;
- the term in (B.37) corresponds to pairs of points from the same family, falling in A_1 and A_2, respectively.

Using (B.23) and (B.33), we observe that (B.34)–(B.36) simplify to

$$\rho_G^2|A_1||A_2| + \sum_{m=0}^{\infty} \rho_Z^2 \int_{A_1}\int_{A_2} b f_Z * \tilde{f}_Z \\ * \sum_{k_1=0}^{m}\sum_{k_2=0}^{m} \binom{m}{k_1}\binom{m}{k_2} q^{2m-k_1-k_2}(\beta p)^{k_1+k_2} f^{*k_1} * \tilde{f}^{*k_2}(y_1 - y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 \tag{B.38}$$

and the term in (B.37) is equal to

$$\rho_Z \sum_{m=1}^{\infty} \iiiint_{A_1}\int_{A_2} \big((\beta p f + q\delta_0)^{*i}(y - x) \\ \cdot [c(\beta p)^2 f(\tilde{y}_1 - y) f(\tilde{y}_2 - y) \\ + \beta p q \left\{ f(\tilde{y}_1 - y)\delta_0(\tilde{y_2} - y) + \delta_0(\tilde{y}_1 - y) f(\tilde{y_2} - y)\right\}] \\ \cdot (\beta p f + q\delta_0)^{*(m-1-i)}(y_1 - \tilde{y}_1) \\ \cdot (\beta p f + q\delta_0)^{*(m-1-i)}(y_2 - \tilde{y}_2)\big)\,\mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}\tilde{y}_1\,\mathrm{d}\tilde{y}_2\,\mathrm{d}y\,\mathrm{d}x. \tag{B.39}$$

In (B.39), y corresponds to an i-th generation point in the family initiated by $x \in Z_{-m}$, $c(\beta p)^2 + 2\beta p q$ is the expected number of pairs of points $\tilde{y}_1$ and $\tilde{y}_2$ which are children of y, and y_1 and y_2 are the $(m-1-i)$-th generation offspring of $\tilde{y}_1$ and $\tilde{y}_2$, respectively. Using Fubini's theorem together with (B.23), after straight forward calculations, (B.39) reduces to

$$\rho_G \int_{A_1}\int_{A_2} \sum_{i=0}^{\infty} \left\{ c(\beta p)^2 f * \tilde{f} + \beta p q (f + \tilde{f}) \right\} \\ * \sum_{k_1=0}^{i}\sum_{k_2=0}^{i} \binom{i}{k_1}\binom{i}{k_2} q^{2i-k_1-k_2}(\beta p)^{k_1+k_2} f^{*k_1} * \tilde{f}^{*k_2}(y_1 - y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2$$

Combining this result with (B.38) we finally see that G_0^{st} has PCF g_G as given by (B.26).

C Simulating the limiting process

This appendix presents an approximate simulation procedure for simulating a special case of G_0^{st} on a bounded region $R \subset \mathbb{R}^d$. It is available in `R`

through the package `icpp`, which can be obtained at `https://github.com/adchSTATS/icpp`. The implementation utilizes existing functions from the packages `spatstat` and `RandomFields` to simulate the noise process.

For simplicity and specificity we make the following assumptions. Let the situation be as in Theorem 4.2, but with $q = 0$ and let $f \sim N_d(\sigma^2)$ with $\sigma > 0$. Also, without loss of generality, assume no thinning (i.e., $p = 1$). Let $R_{\oplus r} = \{\xi \in \mathbb{R}^d : b(\xi, r) \cap R \neq \emptyset\}$ where $b(\xi, r)$ is a closed ball with centre ξ and radius $r \geq 0$. Denote n the number of iterations in our approximate simulation algorithm, that is, $-n$ is the starting time when ignoring what happens previously. Note that $\sqrt{n}\sigma$ is the standard deviation of the nth convolution power of f. To account for edge effects, let $r = 4\sqrt{n}\sigma$ where 4 is an arbitrary non-negative value ensuring that a point of $G^{\mathrm{st}}_{-n} \setminus R_{\oplus r}$ would generate a nth generation offspring in R with very low probability, at most $1/15787$. In the approximate simulation procedure, we ignore those points of $G^{\mathrm{st}}_0 \cap R$ which are generated by an ith generation ancestor x when $i < -n$ or both $-n \leq i < 0$ and $x \notin R_{\oplus 4\sqrt{-i}\sigma}$. This is our algorithm in pseudocode where "parallel-for" means a parallel for loop:

---

parallel-for $i = -n$ to 0 **do**
  simulate $Z'_i := Z_i \cap R_{\oplus 4\sqrt{-i}\sigma}$
end parallel-for
set $O := Z'_{-n}$
if $n \neq 0$ **then**
  for $i = -(n-1)$ to 0 **do**
    parallel-for $x \in O$ **do**
      simulate the 1st generation offspring process, O_x, with parent x
    end parallel-for
    set $O := Z'_i \bigcup \left(\bigcup_{x \in O} O_x \cap R_{\oplus 4\sqrt{-i}\sigma} \right)$
  end for
end if
return O

Note that $\rho_Z \sum_{i=0}^{n} (\beta p)^i$ is the intensity of the stationary point process obtained by ignoring those points of G^{st}_0 which are generated by an ith generation ancestor with $i < -n$. We base the choice of n on this fact by considering a precision parameter $\varepsilon > 0$ and letting

$$n = \sup\left\{ m \in \{1, 2, \ldots\} : \left\| \rho_Z \sum_{i=0}^{m} (\beta p)^i - \rho_G \right\| \leq \varepsilon \right\}.$$

To exemplify, let $\rho_G = 100$ and $\beta p = 0.8$ implying that $\rho_Z = 20$, and let $\varepsilon = 2.22 \times 10^{-16}$, then $n = 159$. If instead $\beta p = 0.99$, then $n = 3609$.

References

Andersen, I. T., Hahn, U., Christensen, E. A., Nejsum, L. N., and Jensen, E. B. V. (2018). Double Cox cluster processes - with applications to photoactivated localization microscopy. *Spatial Statistics*, 27:58–73.

Barndorff-Nielsen, O. E., Kent, J., and Sørensen, M. (1982). Normal variance-mean mixtures and z-distributions. *International Statistical Review*, 50:145–159.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer-Verlag, New York, second edition.

Felsenstein, J. (1975). A pain in the torus: Some difficulties with models of isolation by distance. *The American Naturalist*, 109:359–368.

Kingman, J. F. C. (1977). Remarks on the spatial distribution of a reproducing population. *Journal of Applied Probability*, 14:577–583.

Lavancier, F., Møller, J., and Rubak, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77:853–877.

van Lieshout, M. N. M. and Baddeley, A. (2002). Extrapolating and interpolating spatial patterns. In Lawson, A. and Denison, D., editors, *Spatial Cluster Modelling*, pages 61–86. Chapman & Hall/CRC, Boca Raton, Florida.

Macchi, O. (1975). The coincidence approach to stochastic point processes. *Advances in Applied Probability*, 7:83–122.

Matérn, B. (1960). *Spatial Variation*, Meddelanden från Statens Skogforskningsinstitut, **49** (5).

Matérn, B. (1986). *Spatial Variation*. Lecture Notes in Statistics 36. Springer-Verlag, Berlin.

McCullagh, P. and Møller, J. (2006). The permanental process. *Advances in Applied Probability*, 38:873–888.

Møller, J. (1989). Random tessellations in $\mathbb{R}^d$. *Advances in Applied Probability*, 21:37–73.

Møller, J. (1994). *Lectures on Random Voronoi Tessellations*. Lecture Notes in Statistics 87. Springer-Verlag, New York.

Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability*, 35:614–640.

Møller, J. and Torrisi, G. L. (2005). Generalised shot noise Cox processes. *Advances in Applied Probability*, 37:48–74.

Møller, J. and Torrisi, G. L. (2007). The pair correlation function of spatial hawkes processes. *Statistics and Probability Letters*, 77:995–1003.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, Boca Raton, Florida.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

Neyman, J. and Scott, E. L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 20:1–43.

Shimatani, I. K. (2010). Spatially explicit neutral models for population genetics and community ecology: Extensions of the Neyman-Scott clustering process. *Theoretical Population Biology*, 77:32–41.

Shirai, T. and Takahashi, Y. (2003). Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. *Journal of Functional Analysis*, 205:414–463.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, 36:18–25.

Wiegand, T., Gunatilleke, S., Gumatilleke, N., and Okuda, T. (2007). Analyzing the spatial structures of a Sri Lankan tree species with multiple scales of clustering. *Ecology*, 88:3088–3102.

Paper C

Modelling columnarity of pyramidal cells in the human cerebral cortex

Andreas D. Christoffersen, Jesper Møller, and Heidi S. Christensen

The paper is submitted to the
Journal of the Royal Statistical Society: Series C (Applied Statistics).

The layout has been revised.

Abstract

For modelling the location of pyramidal cells in the human cerebral cortex we suggest a hierarchical point process in $\mathbb{R}^3$. The model consists first of a generalised shot noise Cox process in the xy-plane, providing cylindrical clusters, and next of a Markov random field model on the z-axis, providing either repulsion, aggregation, or both within specified areas of interaction. Several cases of these hierarchical point processes are fitted to two pyramidal cell datasets, and of these a model allowing for both repulsion and attraction between the points seem adequate.

1 Introduction and conclusions

The structuring of neurons in the human brain is a subject of great interest since abnormal structures may be linked to certain neurological diseases (see Casanova, 2007; Esiri and Chance, 2006; Casanova et al., 2006; Buxhoeveden and Casanova, 2002). A specific structure that has been extensively studied in the biological literature is the so called 'minicolumn' structure of the cells in the cerebral cortex (see Buxhoeveden and Casanova, 2002; Rafati et al., 2016, and references therein). Rafati et al. (2016) characterised these minicolumns as 'linear aggregates of neurons organised vertically in units that traverse the cortical layer II–VI, and have in humans a diameter of 35–60 µm and consist typically of 80–100 neurons'.

1.1 Data

In this paper we analyse the structuring of pyramidal cells, which make up approximately 75 % to 80 % of all neurons (Buxhoeveden and Casanova, 2002) and are pyramid shaped cells, where the apical dendrite extends from the top of the pyramid. Specifically, the paper is concerned with two datasets consisting of the locations and orientations of pyramidal cells in a section of the third, respectively, fifth layer of Brodmann's fourth area of the human cerebral cortex. Here, each location is a three-dimensional coordinate representing the centre of a pyramidal cell's nucleolus and each orientation is a unit vector representing the apical dendrite's position relative to the corresponding nucleolus.

Figure C.1 shows the two point pattern datasets of 634 and 548 nucleolus locations which will be referred to as `L3` and `L5`, respectively (for plot of the orientations for `L3`, see Møller et al., 2019). Note that the observation window W for the cell locations is a rectangular region with side lengths 492.70 µm, 132.03 µm, and 407.70 µm for `L3` and 488.40 µm, 138.33 µm, and 495.40 µm for `L5`. Notice also that the nucleolus locations are recorded such that the z-axis is perpendicular to the pial surface of the brain. In accordance to the

minicolumn hypothesis, this implies that the minicolumns extend parallel to the z-axis.

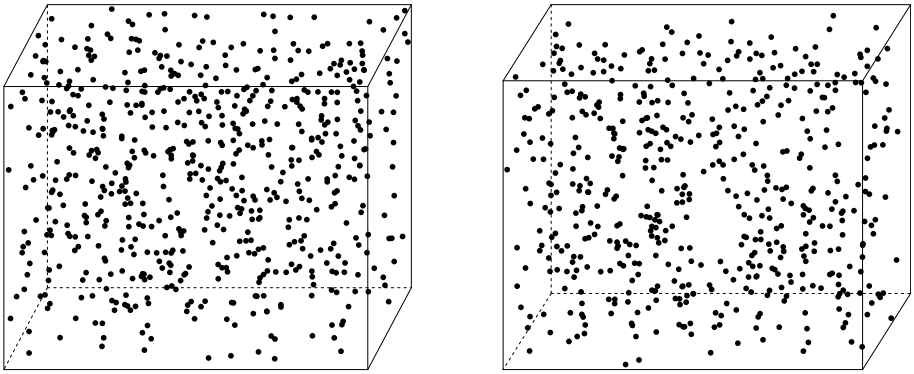

Fig. C.1: Visualisations of the nucleolus locations for datasets `L3` (left) and `L5` (right).

1.2 Background and purpose

Møller et al. (2019) found independence between locations and orientations for `L3` meaning that the two components may be modelled separately; the same conclusion has afterwards been drawn for `L5`. As they also found a suitable inhomogeneous Poisson process model for the orientations, and since it is hard by eye to see much structure in the point patterns shown in Figure C.1, the focus of this paper is on modelling the nucleolus locations. In particular, we aim at modelling the nucleolus locations for `L3` respective `L5` by a spatial point process with a columnar structure and discuss to what extent this relates to the minicolumn hypothesis. Note that for the two datasets we use the same notation X for the spatial point process, and we view X as a random finite subset of W.

To the best of our knowledge the so-called Poisson line cluster point process (see Møller et al., 2016) is the only existing point process model for modelling columnar structures. This model was considered by Rafati et al. (2016) in connection to the pyramidal nucleolus locations, but was not fitted to data. For each point pattern considered in the present paper, we notice later that a more advanced model than the Poisson line cluster point process is needed; below we describe such a model for X.

1.3 Hierarchical point process models

Briefly, we consider a hierarchical model for X (further details are given in Sections 3–5), noting that the observation window is a product space, $W = W_{xy} \times W_z$, where W_{xy} is a rectangular region in the xy-plane and W_z is an

interval on the z-axis. First, we model the point process X_{xy} given by the projection of X onto the xy-plane; second, conditioned on X_{xy}, we model the vector X_z consisting of the z-coordinates of the points in X. Note that the dimension of X_z agrees with the number of points in X_{xy} and is denoted by n.

The model for X_{xy}

For X_{xy} we consider the restriction of a cluster point process to W_{xy} defined briefly as follows (further details are given in Sections 4–5). Let $\Phi \subset \mathbb{R}^2$ be a stationary point process with intensity $\kappa > 0$, and associate to each point $(\xi, \eta) \in \Phi$ a point process $X_{(\xi,\eta)} \subset \mathbb{R}^3$ that is concentrated around the line in $\mathbb{R}^3$ which is perpendicular to the xy-plane, with intersection point $(\xi, \eta, 0)$. We refer to $X_{(\xi,\eta)}$ as the cylindrical cluster associated to (ξ, η). Let $P_{xy}(X_{(\xi,\eta)} \cap W)$ denote the projection onto the xy-plane of the observed part of the cylindrical cluster. For short we refer to the non-empty $P_{xy}(X_{(\xi,\eta)} \cap W)$ as the projected cluster with centre point (ξ, η). Then we let

$$X_{xy} = \bigcup_{(\xi,\eta)\in\Phi} P_{xy}(X_{(\xi,\eta)} \cap W).$$

Further, conditional on Φ, we assume that the projected clusters are independent and each non-empty $P_{xy}(X_{(\xi,\eta)} \cap W)$ is distributed as the intersection of W_{xy} with a finite planar Poisson process translated by the centre point (ξ, η); this Poisson process has intensity function $a\alpha f$, where a is the length of the interval W_z, $\alpha > 0$ is a parameter, and f is the probability density function of a bivariate zero-mean isotropic normal distribution with standard deviation $\sigma > 0$. Thus, ignoring boundary effects, αa is the expected size (or number of points) of a projected cluster and σ controls the spread of points in a projected cluster. Specifically, we let first Φ be a planar stationary Poisson process and second a stationary determinantal point process (Lavancier et al., 2015), since we observe in the first case a very low expected number of points in a projected cluster and because in the second case we want a repulsive model in order to obtain less overlap between the projected clusters.

The special case with Φ a planar stationary Poisson process and X_z a homogeneous binomial point process (that is, the n points in X_z are independent and uniformly distributed on W_z) which is independent of X_{xy} corresponds to a degenerate case of a Poisson line cluster point process as considered in Møller et al. (2016). This becomes clear in Section 4.

The model for X_z conditioned on X_{xy}

We consider several other cases than a homogeneous binomial point process for X_z which is independent of X_{xy}. In general, conditioned on $X_{xy} =$

$\{(x_i, y_i)\}_{i=1}^n$, we propose a Markov random field model, where the conditional probability density function of X_z is of the form

$$f((z_i)_{i=1}^n \mid (x_i, y_i)_{i=1}^n) \propto \gamma_1^{s_{B_1,\theta_1}((z_i)_{i=1}^n \mid (x_i,y_i)_{i=1}^n)} \gamma_2^{s_{B_2,\theta_2}((z_i)_{i=1}^n \mid (x_i,y_i)_{i=1}^n)} \times \mathbb{I}(\|(x_i, y_i, z_i) - (x_j, y_j, z_j)\| > h \text{ for } 1 \leq i < j \leq n), \tag{C.1}$$

with notation defined as follows. We consider $\{(x_i, y_i, z_i)\}_{i=1}^n$ as a realisation of X, where (x_i, y_i) is the xy-point associated to z_i, the realisation of the i'th point in X_z (as a technical detail, unless X_z is a binomial point process, (C.1) is not invariant under permutations of $z_1, \ldots, z_n$ since we have associated (x_i, y_i) to z_i, so we cannot view (C.1) as the density of a point process where we are conditioning on the number of points). Note that the right hand side in (C.1) is an unnormalised density and e.g. $(z_i)_{i=1}^n$ is short hand notation for $(z_1, \ldots, z_n)$. We let $\mathbb{I}(\cdot)$ be the indicator function. Further, $\gamma_1 > 0$, $\gamma_2 > 0$, and $h \geq 0$ are unknown parameters; if $h > 0$, it is a hard core parameter ensuring a minimum distance h between all pair of points in X; for the pyramidal cell data it seems natural to include a hard core condition since cells cannot overlap; and when $\gamma_1 = \gamma_2 = 1$ and $h = 0$, the conditional model simply reduces to the homogeneous binomial point process. Furthermore, for $k = 1, 2$,

$$s_{B_k,\theta_k}((z_i)_{i=1}^n \mid (x_i, y_i)_{i=1}^n) = \sum_{1 \leq i < j \leq n} \mathbb{I}((x_i, y_i, z_i) \in B_k(x_j, y_j, z_j; \theta_k)),$$

where $B_k(x, y, z; \theta_k) \subset \mathbb{R}^3$ is an interaction region, with centre of mass (x, y, z) and a 'size and shape parameter' θ_k, that determines the interaction between points. It is additionally assumed that the hard core ball, given by the three-dimensional closed ball of radius h and centre (x, y, z) does not contain $B_1(x, y, z; \theta_1)$ or $B_2(x, y, z; \theta_2)$. Finally, it is assumed that the symmetry condition

$$(x_i, y_i, z_i) \in B_k(x_j, y_j, z_j; \theta_k) \quad \text{if and only if} \quad (x_j, y_j, z_j) \in B_k(x_i, y_i, z_i; \theta_k)$$

and the disjointness condition

$$B_1(x, y, z; \theta_1) \cap B_2(x, y, z; \theta_2) = \emptyset$$

are satisfied.

These conditions ensure that we can view X_z conditioned on X_{xy} as a Markov random field with second order interactions: for $1 \leq i < j \leq n$, two z-coordinates z_i and z_j interact (in Markov random field terminology, z_i and z_j are neighbours) if and only if $\|(x_i, y_i, z_i) - (x_j, y_j, z_j)\| \leq h$ (that is, the hard core condition is not satisfied, which happens with probability 0) or (x_i, y_i, z_i) lies within the region of interaction of z_j given by the union of $B_1(x_j, y_j, z_j; \theta_1)$ and $B_2(x_j, y_j, z_j; \theta_2)$ (here the symmetry condition is needed

to ensure that we can interchange the roles of i and j). The interaction can either cause repulsion/inhibition or attraction/clumping of the points in X depending on whether $\gamma_k < 1$ or $\gamma_k > 1$ for $k = 1, 2$. Thus, apart from the hard core condition, the model allows for both repulsion and attraction but within different interaction regions B_1 and B_2.

The final hierarchical model and results

At the end of the paper (Section 5) we obtain a satisfactory fit of the following hierarchical model, with the following interpretation of the estimated parameters.

First, the model for X_{xy} is given as in Section 1.3 where the centre process Φ is a most repulsive determinantal point process (as detailed in Section 5.1). The parameter estimates are given in Table C.1, where the estimated expected cluster size $\widehat{\alpha a}$ is much smaller than expected for a minicolumn when restricting it to the observation window – provided the minicolumn hypothesis is true; cf. personal communication with Jens R. Nyengaard. So we neither claim that we have a fitted model for minicolumns nor that the minicolumn hypothesis is true. Instead we have fitted a model with cylindrical clusters: from Table C.1 we see, if $|W_{xy}|$ denotes the area of W_{xy}, the estimated number of projected clusters is $|W_{xy}|\hat{\kappa}$, which is approximately 260 for `L3` and 142 for `L5`; the estimated expected size of a projected cluster is only 2.42 for `L3` and 3.87 for `L5`.

	$\hat{\kappa}$	$\hat{\sigma}$	$\widehat{\alpha a}$
`L3`	0.0040	5.45	2.42
`L5`	0.0021	6.53	3.87

Table C.1: Minimum contrast estimates for our final model of X_{xy} (the DLCPP model in Section 5.1) for the datasets `L3` and `L5`.

Second, the model of X_z conditioned on X_{xy} has cylindrical interaction regions as illustrated in Figure C.2, and (C.1) is the pairwise interaction Markov random field density

$$f((z_i)_{i=1}^n \mid (x_i, y_i)_{i=1}^n) \propto \prod_{1 \le i < j \le n} \mathbb{I}(\|(x_i, y_i, z_i) - (x_j, y_j, z_j)\| > h) \times \gamma_1^{\mathbb{I}(\|(x_i,y_i)-(x_j,y_j)\| \le r_1,\, |z_i - z_j| \le t_1)} \times \gamma_2^{\mathbb{I}(\|(x_i,y_i)-(x_j,y_j)\| \le r_2,\, t_1 < |z_i - z_j| \le t_2)},$$

where $\gamma_1 > 0$ and $\gamma_2 > 0$ are interaction parameters and $0 < r_2 \le r_1$ and $0 < t_1 < t_2$ are parameters which determine the 'range of interaction' such that $h < \sqrt{t_k^2 + r_k^2}$ for $k = 1, 2$. The restrictions on r_1, r_2, t_1, and t_2 are empirically motivated by use of functional summaries as detailed in Section 5.2.

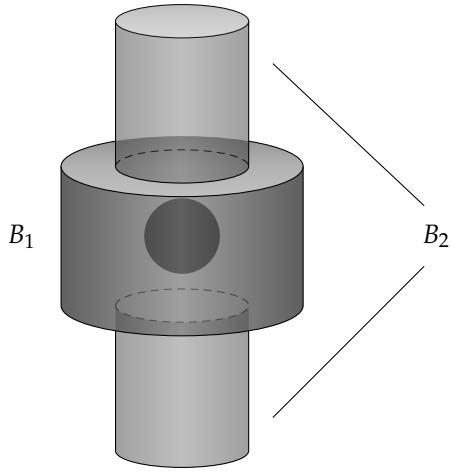

Fig. C.2: Visualisation of the hard core region ball (in dark) and the cylindrical interaction regions B_1 (the cylinder) and B_2 (the union of the two elongated cylinders) used in our final model for `L3`.

The final fitted model have parameter estimates as displayed in Table C.2 where most notably $\hat{\gamma}_1 < 1$ and $\hat{\gamma}_2 > 1$. In particular the final fitted model is in accordance to the empirical findings as noted later when the so-called cylindrical K-function of Figure C.3 is discussed: we have modelled repulsion within stunted cylinders (corresponding to B_1) and aggregation within elongated cylinders (corresponding to B_2), see again Figure C.2. Moreover, the estimated hard core $\hat{h}$ is greater than 6 µm, which is in accordance with 'distance between the nucleolus and the membrane of a pyramidal cell' (personal communication with Jens R. Nyengaard). Note that the hard core ball is much smaller than the interaction region B_1: $2\hat{h}$ (the diameter of the hard core ball) is about half as small as $2\hat{t}_1$ (the height of B_1). Finally, comparing Tables C.1-C.2, we note that the two 'clustering parameters' $2\hat{\sigma}$ and $\hat{r}_2$ are of the same order.

	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{h}$	$\hat{r}_1$	$\hat{t}_1$	$\hat{r}_2$	$\hat{t}_2$
`L3`	0.41	1.78	6.25	20	11.5	11	35.5
`L5`	0.51	1.68	6.77	24.25	15.5	14.75	37.25

Table C.2: Pseudo likelihood estimates of our final model (model 5 from Table C.4 in Section 5.2) for the datasets `L3` and `L5`.

In conclusion, for each dataset we have fitted a rather complex hierarchical point process model describing columnar structures of the nucleolus locations. This model included repulsion between nucleolus locations given by a hard core condition on a small scale and a stunted cylindrical interac-

tion region on a larger scale, as well as clustering between nucleolus locations given by an elongated cylindrical interaction region.

1.4 Model fitting

In Møller et al. (2016) parameter estimation for the degenerate PLCPP model was simply done by a moment based procedure which included minimisation of a certain contrast between a theoretical second order moment functional summary and its empirical estimate. In the present paper we use a similar minimum contrast procedure for estimating the parameters of models for X_{xy}. For the models of X_z conditioned on X_{xy} we find it convenient to use a maximum pseudo likelihood procedure as detailed in Section 5.2. Moreover, each fitted model is evaluated by considering informative global extreme rank length (GERL) envelope procedures (Mrkvička et al., 2018; Myllymäki et al., 2017) for various functional summaries.

1.5 Outline

The remainder of this paper explains how we arrive at the final model given in Section 1.3 after fitting several other models. In Section 2 we introduce some basic concepts and definitions needed for the models in the subsequent sections. In Section 3 we investigate how the nucleolus locations deviate from complete spatial randomness (that is, when X is a homogeneous Poisson process), and in Section 4 we also notice a deviation from a fitted degenerate PLCPP model. Finally, in Section 5 we introduce and fit various generalisations of the degenerate PLCPP model as briefly described in Sections 1.3–1.4.

2 Preliminaries

The point processes X, X_{xy}, and X_z introduced above are viewed as the restriction to the bounded sets W, W_{xy}, and W_z of a locally finite point process $Y \subset \mathbb{R}^d$ with $d = 3, 2, 1$, respectively. Briefly speaking, this means that Y is a random subset of $\mathbb{R}^d$ satisfying that $Y_B = Y \cap B$ is finite for any bounded set $B \subset \mathbb{R}^3$; for a more rigorous definition of point processes, see e.g. Daley and Vere-Jones (2003) or Møller and Waagepetersen (2004). Below we recall a few basic statistical tools needed in this paper, using the generic notation Y for a locally finite point process defined on $\mathbb{R}^d$ (apart from the cases above, we have in mind that Y could also be the centre process Φ from Section 1.3).

2.1 Moments

For each integer $n \geq 1$, to describe the n'th order moment properties of Y, we consider the so-called n'th order intensity function $\lambda^{(n)} : (\mathbb{R}^d)^n \to [0, \infty)$

given that it exists. This means that for any pairwise distinct and bounded Borel sets $B_1, \ldots, B_n \subset \mathbb{R}^d$,

$$\mathrm{E}\left[n(Y_{B_1}) \cdots n(Y_{B_n})\right] = \int_{B_1} \cdots \int_{B_n} \lambda^{(n)}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

is finite, where $n(Y_B)$ denotes the cardinality of Y_B.

The first order intensity function $\lambda^{(1)} = \lambda$ is of particular interest and is simply referred to as the intensity function. Heuristically, $\lambda(u)\,\mathrm{d}u$ can be interpreted as the probability of observing a point from Y in the infinitesimal ball of volume $\mathrm{d}u$ centred at u. If the intensity function $\lambda(\cdot) \equiv \lambda$ is constant, then $\lambda|B| = \mathrm{E}\left[n(Y_B)\right]$ for any bounded Borel set $B \subset \mathbb{R}^d$, where $|\cdot|$ is the Lebesgue measure. In this case Y is said to be homogeneous and otherwise inhomogeneous. Clearly, stationarity of Y (meaning that its distribution is invariant under translations in $\mathbb{R}^d$) implies homogeneity.

2.2 Functional summaries

In order to determine an appropriate model for an observed point pattern, we consider functional summaries, which reflect/summarise different properties of the point pattern and are useful for model fitting and control. The main examples are considered below.

To summarise the second order moment properties, it is custom to consider the pair correlation function (PCF), g, which is defined as the ratio of the second and first order intensity function, that is,

$$g(x_1, x_2) = \frac{\lambda^{(2)}(x_1, x_2)}{\lambda(x_1)\lambda(x_2)}, \quad x_1, x_2 \in \mathbb{R}^d.$$

Heuristically, $g(x_1, x_2)$ can be interpreted as the probability of simultaneously observing a point from X in each of the two infinitesimal balls of volume $\mathrm{d}x_1$ and $\mathrm{d}x_2$ centred at respectively x_1 and x_2 relative to the probability of independently observing a point in the two infinitesimal balls. The PCF is said to be stationary when (with abuse of notation) $g(x_1, x_2) = g(x_1 - x_2)$ and isotropic when $g(x_1, x_2) = g(\|x_1 - x_2\|)$.

If the PCF is stationary, it is closely related to the so-called second order reduced moment measure, $\mathcal{K}$, given by

$$\mathcal{K}(B) = \int_B g(x)\,\mathrm{d}x,$$

where $B \subset \mathbb{R}^d$ is a Borel set (see Møller and Waagepetersen, 2004). If Y is stationary and B has centre of mass at the origin of $\mathbb{R}^d$, then $\lambda\mathcal{K}(B)$ can be interpreted as the expected number of further points falling within B given that Y has a point at the origin; and when considering scalings of B, we refer

to B as a structuring element. The simplest example occurs when B is a ball centred at the origin and with radius $r > 0$; then $K(r) = \mathcal{K}(B)$ becomes the K-function introduced by Ripley (1976); and often we instead consider a transformation of the K-function, which is called the L-function and defined by $L(r) = (K(r)/\omega_d)^{1/d}$, where ω_d is the volume of the d-dimensional unit ball. In particular, if Y is a stationary Poisson process, then $L(r) = r$.

For detecting cylindrical structures, Møller et al. (2016) introduced the cylindrical K-function which corresponds to $\mathcal{K}(B)$ when B is a cylinder of height $2t$, base-radius r, and centre of mass at the origin. Note that Ripley's K-function depends only on one argument, r, while the cylindrical K-function depends both on r, t, and the direction of the cylinder. However, when $d = 3$ and since the minicolumns are expected to extend along the z-axis, we only consider cylinders extending in this direction, effectively reducing the number of arguments to two.

We will also consider the commonly used F-, G-, and J-functions when performing model control; see van Lieshout and Baddeley (1996) for definitions. Briefly, if Y is stationary, $F(r)$ is the probability that Y has a point within distance $r > 0$ from a fixed location in $\mathbb{R}^d$; $G(r)$ is the probability that Y has another point within distance $r > 0$ from an arbitrary fixed point in Y; and $J(r) = (1 - G(r))/(1 - F(r))$ when $F(r) < 1$.

The functional summaries will in the following be used both for model fitting as described in Section 2.3 and for model checking using GERL envelope procedures as mentioned in Section 1.4. In the GERL envelope procedure, the distribution of the empirical functional summary under the hypothesis of interest is estimated by simulations. The procedure is a refinement of the global rank envelope procedure (Myllymäki et al., 2017), where it is recommended to use 2499 simulations for a single one-dimensional functional summary and at least 9999 simulations for a single two-dimensional functional summary (Mrkvička et al., 2016). However, we consider a concatenation of the L-, G-, F-, and J-functions, as well as the cylindrical K-function in which case Mrkvička et al. (2017) recommend using more simulations. Particularly for a concatenation of k one-dimensional summary functions they recommend using $k \times 2499$ simulations. We do however have a different setup since we are concatenating both one- and two-dimensional summary functions. For the GERL envelope procedure, Mrkvička et al. (2018) suggest that a lower number of simulations may be enough. Therefore, we use 9999 simulations. Since we consider a concatenation of one- and two-dimensional functional summaries, we ensure that each of the functional summaries are weighted equally in the GERL envelope test by evaluating them at the same number of arguments (Mrkvička et al., 2017). Specifically we consider 64^2 r-values for each of the L-, G- F-, and J-functions and a square grid over 64 r-values and 64 t-values for the cylindrical K-function.

2.3 Minimum contrast estimation

For parametric point process models, minimum contrast estimation is a computationally simple fitting procedure introduced by Diggle and Gratton (1984) that is applicable when a closed form expression of a functional summary, T, exists. The idea is to minimise the distance from the theoretical function T to its empirical estimate $\hat{T}$ for the data. Specifically, if T depends on the parameter vector θ and is a function of 'distance' $r > 0$ (as for example in case of Ripley's K-function), the minimum contrast estimate of θ is given by

$$\hat{\theta} = \operatorname{argmin}_{\theta} \int_{r_{\min}}^{r_{\max}} \left| T(\theta, r)^q - \hat{T}(r)^q \right|^p \mathrm{d}r, \tag{C.2}$$

where $r_{\min} < r_{\max}$, q, and p are positive tuning parameters. General recommendations on q are given in Guan (2009) and Diggle (2014), when $T(r) = g(r)$ or $T(r) = K(r)$. Unless otherwise stated, we let $p = 2$, $q = 1/4$, $r_{\min} = 0$, and $r_{\max}$ be one fourth of the shortest side length of the relevant observation window (the rectangular window W_{xy} in our case).

When the PCF has a closed form expression, alternative estimation procedures can be used, including the second order composite likelihood (see Guan, 2006; Waagepetersen, 2007), adapted second order composite likelihood (see Lavancier et al., 2018), and Palm likelihood (see Ogata and Katsura, 1991; Prokešová et al., 2016; Baddeley et al., 2015).

3 Complete spatial randomness

The most natural place to begin our point pattern analysis is by testing whether a homogeneous Poisson process X with intensity $\lambda > 0$ (we then view Y as a stationary Poisson process with the same intensity), also called complete spatial randomness (CSR), adequately describe each nucleolus point pattern dataset. Recall that this means that $n(X)$ is Poisson distributed with parameter $\lambda|W|$ and conditional on $n(X)$ the points in X are independent and uniformly distributed within W. Even when CSR is not an appropriate model, deviations from the model can be useful for determining whether the points of a homogeneous point pattern tend to e.g. attract or repel each other.

The CSR model is fully specified by its intensity, which naturally is estimated by $n(X)/|W|$, which is equal to 2.37×10^{-5} for `L3` and 1.63×10^{-5} for `L5`. For this fitted model Figure C.3 summarises the results of the GERL envelope procedure based on the concatenation of the L-, G-, F-, J-, and cylindrical K-functions as discussed in Section 2.2. Particularly, the left column depicts the part concerning the empirical functional summaries $\hat{L}(r) - r$, $\hat{G}(r)$, $\hat{F}(r)$, and $\hat{J}(r)$ along with the corresponding 95% envelope. The right column depicts the empirical cylindrical K-function along with the areas at which it falls

outside the 95% envelope. It is observed that the empirical functional summaries $\hat{L}$, $\hat{F}$, and $\hat{J}$ fall strictly outside the envelope for midrange values of r in a manner that indicates repulsion between points at this range. For small and large r-values the observed point patterns resemble the Poisson process. This behaviour could suggest a kind of clustering, where clusters of points from a Poisson process are somewhat separated. The separation of these clusters seems to be more pronounced for `L3` than for `L5`. Further, in the right column of Figure C.3, the empirical cylindrical K-function falls above the upper global rank envelopes for cylinders that have a height larger than approximately 35 µm for both datasets and a base radius of approximately 5 µm to 15 µm for `L3` and 5 µm to 20 µm for `L5`. Furthermore, the observed cylindrical K-functions falls below the lower 95% GERL envelope for cylinders with a height of approximately 10 µm to 30 µm and a base radius larger than 5 µm. Hence, for elongated cylinders extending in the z-direction, we tend to see more points in the data than we expect under CSR, while for stunted cylinders we tend to see fewer points. This seems to be in correspondence with columnar structures where the columns extend in the z-direction.

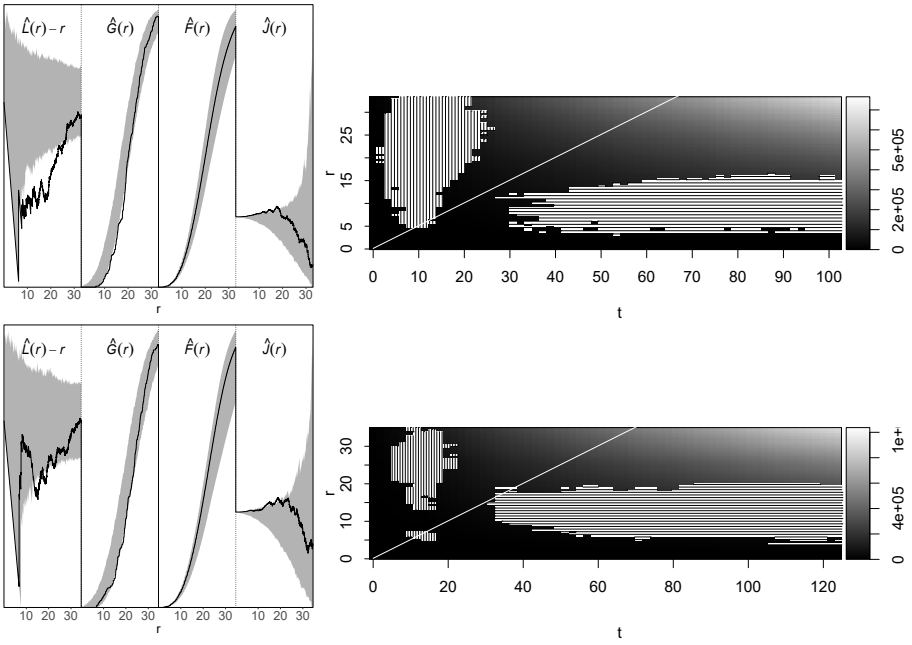

Fig. C.3: Results of the GERL envelope procedure under CSR based on a concatenation of the empirical *L*-, *G*-, *F*-, *J*-, and cylindrical *K*-functions. Left: concatenation of the one-dimensional empirical functional summaries for the data (solid line) together with 95% envelopes (grey region); for ease of visualisation, the functions have been scaled. Right: empirical cylindrical *K*-function (grey scale) where shaded vertical/horizontal lines indicate that the function falls above/below the 95% envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the dataset `L3`. Bottom: results for the dataset `L5`.

4 The degenerate Poisson line cluster point process

Møller et al. (2016) presented the so-called Poisson line cluster point process (PLCPP) which is useful for modelling columnar structures. Specifically, we consider a degenerate PLCPP $Y \subset \mathbb{R}^3$ constructed as follows.

1. Generate a stationary Poisson process $\Phi = \{(\xi_i, \eta_i)\}_{i=1}^{\infty} \subset \mathbb{R}^2$ with finite intensity $\kappa > 0$. Each point $(\xi_i, \eta_i) \in \Phi$ corresponds to an infinite line l_i in $\mathbb{R}^3$ which is perpendicular to the xy-plane, that is, $l_i = \{(\xi_i, \eta_i, z) \,|\, z \in \mathbb{R}\}$.

2. Conditional on Φ, generate independent stationary Poisson processes $L_1 \subset l_1, L_2 \subset l_2, \ldots$ with identical and finite intensity $\alpha > 0$.

3. Generate point processes $X_1, X_2, \ldots \subset \mathbb{R}^3$ by independently displacing the points of $L_1, L_2, \ldots$ by the zero-mean isotropic normal distribution with standard deviation $\sigma > 0$.

4. Finally, set $Y = \bigcup_{i=1}^{\infty} X_i$ and $X = Y_W$.

Some comments to the construction in items 1–4 are in order.

In the general definition of the PLCPP in Møller et al. (2016), the lines $l_1, l_2, \ldots$ are modelled as a stationary Poisson line process. That is, the lines are not required to be perpendicular to the xy-plane nor does the Poisson line process need to be degenerate (meaning that the lines are not required to be mutually parallel). Further, the dispersion density (used in item 3) can be arbitrary. However, the construction is still such that Y becomes stationary. Furthermore, it turns out that it does not matter whether we consider a three-dimensional normal distribution for displacements in in item 3 or a bivariate normal distribution with displacements of the xy-coordinates for the points of $L_1, L_2, \ldots$.

Returning to the degenerate PLCPP of items 1–4, we imagine that each X_i is a cylindrical cluster of points around the line l_i, where these cylindrical clusters are parallel to the z-axis. Furthermore, the interpretation of the parameters κ, α, and σ in terms of a Poisson cluster point process is similar to that in Section 1.3 except that we now also consider lines not intersecting W: if Y as defined by items 1–4 is restricted to a subset $S \subset \mathbb{R}^3$ bounded by two planes parallel to the xy-plane, for specificity $S = \{(x, y, z) \in \mathbb{R}^3 \,|\, z \in W_z\}$, this restricted point process can be seen as a (modified) Thomas process (see Thomas, 1949; Møller and Waagepetersen, 2004) on $\mathbb{R}^2$ along with independent z-coordinates following a uniform distribution on W_z.

To see this, first note that conditional on $\Phi = \{(\xi_i, \eta_i)\}_{i=1}^{\infty}$ and for all $i = 1, 2, \ldots$, X_i is a Poisson process in $\mathbb{R}^3$ with intensity function $\lambda_i((x, y,$

$z)) = \alpha f(x - \xi_i, y - \eta_i)$, where f is the probability density function of the bivariate isotropic normal distribution given in item 3. In turn, this implies that Y conditioned on Φ is a Poisson process in $\mathbb{R}^3$ with intensity function $\sum_{i=1}^{\infty} \lambda_i((x, y, z))$. Further, since $\lambda_i(x, y, z) = \lambda_i(x, y)$ does not depend on z for all $i = 1, 2, \ldots$, the projection of Y_S onto the xy-plane, $P_{xy}(Y_S)$, conditioned on Φ is a Poisson process with intensity $a \sum_{i=1}^{\infty} \lambda_i(x, y)$, where a is the length of the interval W_z. Since Φ is a stationary Poisson process, $P_{xy}(Y_S)$ is a Thomas process with centre process intensity κ and expected cluster size αa (that is, the expected number of points in $X_i \cap S$). Finally, from items 2–4 it follows that the z-coordinates of X_z are independent and uniformly distributed on W_z, and they are independent of X_{xy}.

Consequently, simulating $X = Y_W$ is straightforwardly done by simulating a Thomas point process (on a larger set than W_{xy} in order to avoid boundary effects) along with independent uniform z-coordinates on W_z. For simulating the Thomas point process we apply standard software from the `R`-package `spatstat` (Baddeley et al., 2015). Similarly, fitting a degenerate PLCPP to a realisation of X is simply a matter of fitting a Thomas process to the point pattern consisting of the xy-coordinates of the points in that realisation. Since the K-function of the Thomas process has a closed form expression, the model can be fitted using minimum contrast estimation with $T(r) = K(r)$ in (C.2). Table C.3 summarises the parameter estimates, where most notably the expected cluster size $\widehat{\alpha a}$ is < 1 for both `L3` and `L5`. Understanding each cylindrical cluster within W as (a part of) a minicolumn, 'these parameter estimates result in very unnatural models for the datasets, since each minicolumn within W is expected to consist of less than one point' (personal communication with Jens R. Nyengaard).

	$\hat{\kappa}$	$\hat{\sigma}$	$\widehat{\alpha a}$
`L3`	0.027	2.86	0.36
`L5`	0.0085	4.58	0.95

Table C.3: Minimum contrast estimates of the degenerate PLCPP.

Despite the fact that the fitted degenerate PLCPP models are somewhat unnatural and hardly can be interpreted as a model with (hypothesised) minicolumnar structures, GERL envelope procedure based on a concatenation of the F-, G-, and J-functions show that the Thomas process suitably fit the projected locations with a p-value of 0.76 for `L3` and 0.87 for `L5`. However, results from the concatenated GERL envelope procedure described in Section 2.2 indicated that the model did not suitably describe the three-dimensional nucleolus locations with a p-value of 10^{-4} for both `L3` and `L5`. Specifically, Figure C.4 shows the empirical cylindrical K-function and indicates where it deviates from the 95% envelope. Clearly, the model does

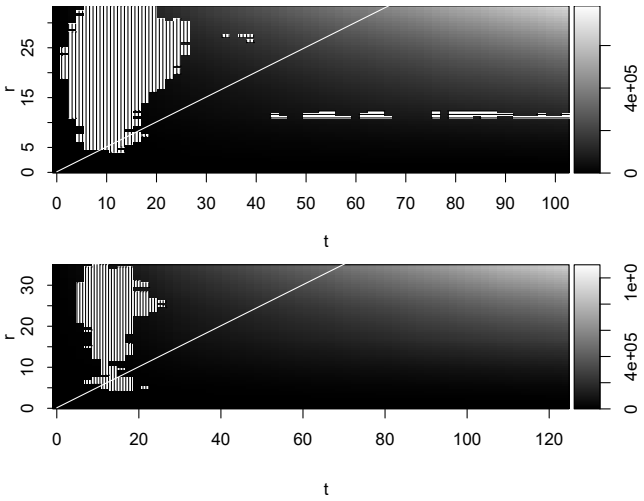

Fig. C.4: Empirical estimates of the cylindrical K-function (grey scale) where shaded vertical/horizontal lines indicate that the function falls above/below the 95% GERL envelope under the fitted degenerate PLCPP and based on the concatenation described in Section 2.2. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the dataset `L3`. Bottom: results for the dataset `L5`.

account for some of the columnarity of the data as opposed to CSR, but the empirical cylindrical K-function for `L3` still falls above the 95% envelope. Furthermore, the empirical cylindrical K-function for both datasets falls below the 95% envelope similar to what was seen under CSR, indicating a lack of regularity, which in fact is supported by the one-dimensional functional summaries (not shown). This could suggest that the cylindrical clusters should be more distinct; motivating us to generalise the degenerate PLCPP model as in the following section.

5 A generalisation of the degenerate PLCPP

As some but not all features of the data were explained by the degenerate PLCPP fitted in Section 4, we propose in this section two generalisations as follows.

1. The centre process Φ is a planar stationary point process.
2. X_z conditioned on X_{xy} follows a Markov random field model.

The first modification is straightforward and for this specific application we chose a repulsive centre process to obtain more distinguishable cylindrical clusters; this is detailed in Section 5.1. Further, the assumption of stationarity of Φ is made in order to apply a similar minimum contrast estimation

procedure as in Section 4, so implicitly we make the assumption that the PCF or the K-function is expressible on closed form. For the second modification we suggest a conditional model inspired by the multiscale point process and particularly the Strauss hard core point process (see e.g. Møller and Waagepetersen, 2004) which will allow for further repulsion or even aggregation between the points; this is detailed in Section 5.2.

5.1 A determinantal point process model for the centre points

Consider a point process $Y \subset \mathbb{R}^3$ specified by items 1–4 in Section 4 with the exception that the centre process Φ now is an arbitrary stationary planar point process. Then, recalling the notation from Section 4, $P_{xy}(Y_S)$ is a planar Cox process (see Møller and Waagepetersen, 2004) and even a planar generalised shot-noise Cox process (see Møller and Torrisi, 2005) driven by the random intensity function $\Lambda(x,y) = a\sum_{i=1}^{\infty}\lambda_i(x,y)$ for $(x,y) \in \mathbb{R}^2$. Moreover, $P_{xy}(Y_S)$ corresponds to the Thomas process, but with a different centre point process (unless of course Φ is a stationary Poisson process).

In this section we focus on the case where Φ is a stationary determinantal point process (DPP; see Lavancier et al., 2015), in which case we will refer to Y as the determinantal line cluster point process (DLCPP). A DPP is defined in terms of its n'th order intensity function for $n = 1, 2, \ldots$: let $C : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{C}$ be a function and $\lambda^{(n)}$ the n'th order intensity function of Φ, then Φ is called a DPP with kernel C if

$$\lambda^{(n)}(x_1, \ldots, x_n) = \det[C](x_1, \ldots, x_n) \qquad \text{for } n = 1, 2, \ldots, x_1, \ldots, x_n \in \mathbb{R}^2,$$

where $\det[C](x_1, \ldots, x_n)$ is the determinant of the $n \times n$ matrix with (i,j)'th entry $C(x_i, x_j)$. For further details on DPPs, we refer to Lavancier et al. (2015) and the references therein. When Φ is a DPP, we call $P_{xy}(Y_S)$ a determinantal Thomas point process (DTPP). The DTPP is discussed to some extent in Møller and Christoffersen (2018), where a closed form expression of its PCF is found. Thus, the DLCPP can be fitted by fitting a DTPP to the projected data using a minimum contrast procedure (see Section 2.3).

For our data we want to obtain a DLCPP with as much repulsion as possible between the centre lines of the cylindrical clusters. Therefore, we let Φ be the 'most repulsive DPP' (in the sense of Lavancier et al., 2015), which is the jinc-like DPP given by the kernel $C(x_1, x_2) = \sqrt{\rho/\pi}J_1\left(2\sqrt{\pi\rho}\|x_1 - x_2\|\right)/\|x_1 - x_2\|$, where J_1 is the first order Bessel function of the first kind and $\|\cdot\|$ denotes the usual planar distance (for more information on this particular DPP, see Lavancier et al., 2018; Biscio and Lavancier, 2016).

Simulation of the DTPP is done by first simulating a DPP with intensity κ (on a larger region than W_{xy} in order to avoid boundary effects), for which we use the functionality of `spatstat`, then secondly generating for each cluster a

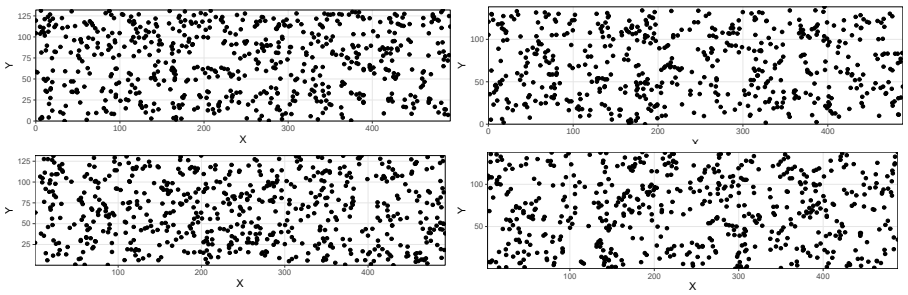

Fig. C.5: Projection of observed nucleolus locations onto the xy-plane (left) and simulations from the fitted jinc-like DTPP (right) for the datasets `L3` (top) and `L5` (bottom).

Poisson distributed number of points with intensity αa, and finally displacing these points by a bivariate zero-mean isotropic normal distribution.

The parameter estimates of the jinc-like DTPP model were obtained by minimum contrast with $T(r) = g(r)$; see Table C.1 for the results and the accompanying discussion in Section 1.3. Despite the expectation under the minicolumn hypothesis of having much higher values of $\widehat{\alpha a}$ than in Table C.1 (see again Section 1.3), simulations of the fitted jinc-like DPP in the xy-plane seem in reasonable correspondence to the projected data; see Figure C.5. Furthermore, results from the GERL envelope procedure based on a concatenation of the F-, G-, and J-functions do not provide any evidence against the jinc-like DPP model for the projected points with p-values of 0.67 for `L3` and 0.83 for `L5`.

Since the jinc-like DTPP model fits the projected data well, we proceeded and added independent uniform z-coordinates on W_z to the simulations, thereby obtaining simulations of the jinc-like DLCPP. Figure C.6 summarises the result of the 95% GERL test based on the concatenation of functional summaries as described in Section 2.2. The left column depicts the part of the one-dimensional functional summaries along with 95% envelopes, while the right column shows the empirical cylindrical K-function along with shaded regions that indicate where it deviates from the corresponding envelope. These plots show that the models do not account for the regularity of the data. This leads us to our next generalisation in Section 5.2.

5.2 A Markov random field model for the z-coordinates

Motivated by the observations at the end of the previous section, in this section we propose to model the z-coordinates conditioned on the xy-coordinates by a pairwise interaction point process as given in (C.1). Thereby, our hierarchical model construction yields a more flexible model for X but we ignore edge effects in the sense that we have only specified a model for first $P_{xy}(Y_S)$

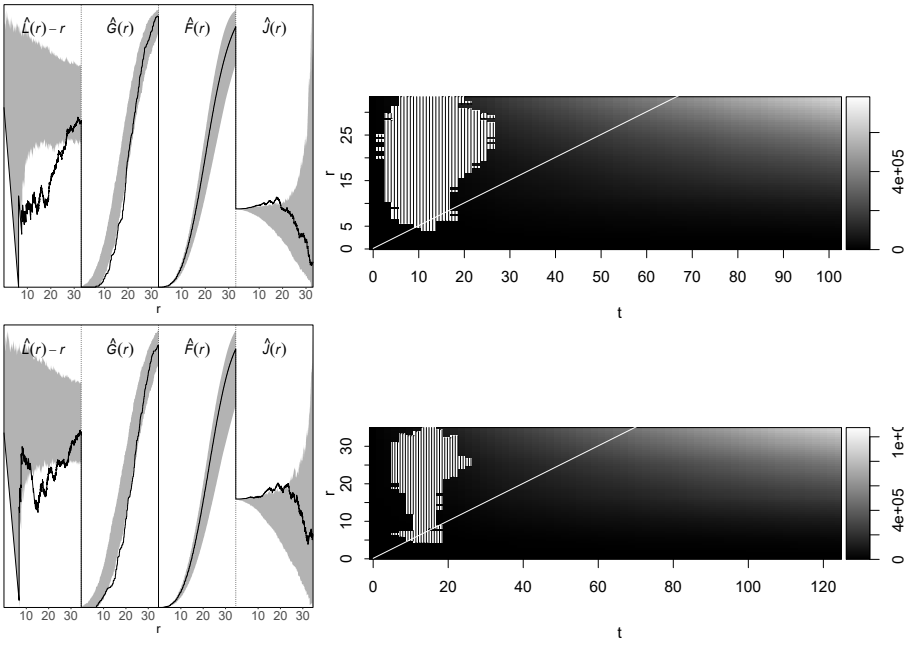

Fig. C.6: Results of the GERL envelope procedure under the fitted DLCPP based on a concatenation of the empirical *L*-, *G*-, *F*-, *J*-, and cylindrical *K*-functions. Left: concatenation of the one-dimensional empirical functional summaries for the data (solid line) together with 95% envelopes (grey region); for ease of visualisation, the functions have been scaled. Right: empirical cylindrical *K*-function (grey scale) where shaded vertical/horizontal lines indicate that the function falls above/below the 95% envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the dataset `L3`. Bottom: results for the dataset `L5`.

and second X_z conditioned on $X_{xy} = P_{xy}(Y_S) \cap W_{xy}$, thereby ignoring a possible influence of points in $Y \setminus W$ when (C.1) is used in the latter step (unless it specifies a binomial point process). This simplification is just made for mathematical convenience; indeed it would be interesting to construct a model taking edge effects into account so that Y becomes stationary, but we leave this challenging issue for future research. Below we first specify the ingredients of the conditional probability density function given in (C.1) for various models and discuss the overall conclusions, next describe how to find parameter estimates, and finally discuss how well the estimated models fit the data. Note that although we have not specified a stationary model for Y, it may still make sense to interpret plots of empirical cylindrical K-functions and $\hat{F}$, $\hat{G}$, $\hat{J}$, and $\hat{L}$-functions, since we have stationarity in the xy-plane and approximately stationarity in the z-direction (as the density (C.1) is invariant under 'translations of $(z_1, \ldots, z_n)$ within W_z').

In our search for a suitable model for the nucleolus locations, we considered many special cases of (C.1). Table C.4 summarises five selected models, where $b((x,y,z);r)$ is the ball with centre (x,y,z) and radius r, and where $c((x,y,z);r,t)$ and $d((x,y,z);r,t)$ denote the cylinder and double cone, respectively, with centre of mass at (x,y,z), height $2t$, base radius r, and extending in the z-direction. First, we considered model 1 which is a hard core model if $h > 0$ and one of the simplest ways of modelling regularity; note that model 1 with $h = 0$ is the binomial point process with a uniform density as considered in Section 4. Though accounting for small distance repulsion, when fitted to the data, model 1 turned out not to account for the repulsion at larger scales. Second, we considered model 2 which is a conditional Strauss model with a hard core condition (see Møller and Waagepetersen, 2004, and the references therein). For this model the scale of repulsion for the z-coordinates seemed too great for points with similar xy-coordinates, and therefore we found it natural to replace the spherical interaction region with a cylinder, yielding model 3. However, model 3 did not correct the problem, and continuing with a single region of interaction we next suggested model 4 with a region given by a cylinder minus a double cone. Model 4 does to a smaller degree penalise the occurrence of points with similar xy-coordinates. However, this model was not suitable either. Models 1–4 were discarded by GERL tests with extremely small p-values. Finally, we considered model 5 which is a more flexible model that allows for both repulsion and aggregation within cylinder shaped interaction regions, cf. the discussion in Section 1.3. For simplicity all the models were also considered without a hard core condition, that is $h = 0$, but was in every case found inadequate.

The likelihood function corresponding to (C.1) involves a normalising constant which needs to be approximated by Markov chain Monte Carlo methods. We propose an easier alternative based on the pseudo likelihood function (Besag, 1975) defined as follows when the data is given by $\{(x_i, y_i,$

Model	γ_1	γ_2	$B_1(\cdot;\theta_1)$	$B_2(\cdot;\theta_2)$	θ_1
1	1	1	$\emptyset$	$\emptyset$	-
2	>0	1	$b(\cdot;r)$	$\emptyset$	$r>h$
3	>0	1	$c(\cdot;r,t)$	$\emptyset$	$r,t>0$
4	>0	1	$c(\cdot;r,t)\backslash d(\cdot;r,t)$	$\emptyset$	$r,t>0$
5	>0	>0	$c(\cdot;r_1,t_1)$	$c(\cdot;r_2,t_2)\setminus c(\cdot;r_1,t_1)$	$r_1,t_1>0$

Table C.4: Specific choices of the parameters $\gamma_1,\gamma_2,\theta_1,\theta_2$ and the interaction regions $B_1(\cdot;\theta_1), B_2(\cdot;\theta_2)$ for five models given by the density (C.1). For each model, a hard core parameter $h \geq 0$ is included. Apart from the specified restrictions, it is required for models 2–5 that $B_1(\cdot;\theta_1) \not\subseteq b(\cdot;h)$ (for model 2 this means that $r > h$ as already indicated) and in addition for model 5 that $B_2(\cdot;\theta_2) \not\subseteq b(\cdot;h)$ where $\theta_2 = (r_2,t_2)$ with $r_1 \geq r_2 > 0$ and $t_2 > t_1$.

$z_i)\}_{i=1}^n \subset W$. For $i = 1,\ldots,n$, the i'th full conditional density associated to (C.1) is

$$\begin{aligned} &f(z_i \mid (z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_n),(x_j,y_j)_{j=1}^n) \\ &\quad = \mathbb{I}(\|(x_i,y_i,z_i)-(x_j,y_j,z_j)\| > h \text{ for } j\neq i)\gamma_1^{s_{1,i}}\gamma_2^{s_{2,i}}/c_i \end{aligned} \tag{C.3}$$

where we define

$$s_{k,i} = \sum_{j:\, j\neq i} \mathbb{I}\left((x_j,y_j,z_j) \in B_k((x_i,y_i,z_i);\theta_k)\right), \qquad k=1,2,$$

and where the normalising constant is given by

$$\begin{aligned} c_i = \sum_{k=0}^{n-1}\sum_{l=0}^{n-1}\gamma_1^k\gamma_2^l \int_{W_z} &\mathbb{I}(\|(x_i,y_i,z)-(x_j,y_j,z_j)\| > h \text{ for } j\neq i) \\ &\times \mathbb{I}\Big(\sum_{j:\, j\neq i}\mathbb{I}\left((x_j,y_j,z_j)\in B_1((x_i,y_i,z);\theta_1)\right)=k\Big) \\ &\times \mathbb{I}\Big(\sum_{j:\, j\neq i}\mathbb{I}\left((x_j,y_j,z_j)\in B_2((x_i,y_i,z);\theta_2)\right)=l\Big)\,\mathrm{d}z. \end{aligned}$$

To estimate the model parameters we maximise the log pseudo likelihood given by

$$\begin{aligned} &LP(\gamma_1,\gamma_2,h,\theta_1,\theta_2) \\ &\quad = \sum_{i=1}^n \log f(z_i \mid (z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_n),(x_j,y_j)_{j=1}^n). \end{aligned} \tag{C.4}$$

Clearly, by (C.3) the maximum pseudo likelihood estimate (MPLE) $\hat{h}$ of h is the minimum distance between any distinct pair of points (x_i,y_i,z_i) and

(x_j, y_j, z_j) in the data. This in fact also corresponds to the maximum likelihood estimate. For $h = \hat{h}$ and for fixed θ_1 and θ_2, we easily obtain the following. For each of models 2–4, the MPLE of γ_1 exists if and only if $s_{1,i} \neq 0$ for some i, and then the log pseudo likelihood function is strictly concave with respect to $\log \gamma_1$. For model 5, the MPLE of (γ_1, γ_2) exists if and only if $s_{1,i} \neq 0$ for some i and $s_{2,j} \neq 0$ for some j, and then the log pseudo likelihood function is strictly concave with respect to $(\log \gamma_1, \log \gamma_2)$. Therefore, the (profile) log pseudo likelihood can be maximised by a combination of a grid search over θ_1 and θ_2 and numerical optimisation with respect to γ_1 and γ_2. Table C.2 shows the maximum pseudo likelihood estimates of model 5 for the two datasets, where for the numerical optimisation we used `optim` (a general-purpose optimisation function from the `R`-package `stats`).

Each of the five models in Table C.4 were fitted to `L3` and `L5` by finding the maximum pseudo likelihood estimate, and model checking was performed using GERL envelope procedures based on the concatenation of functional summaries as discussed in Section 2.2. For the fitted models, model 5 was the most appropriate with p-values of 0.34 for `L3` and 0.03 for `L5` when using the GERL envelope procedure; the 95% GERL envelope is visualised in Figure C.7. Thus no evidence is seen against the fitted models summarised in Table C.2 for `L3` while only slight evidence is present for `L5`. We note that for both datasets the fitted models are such that B_1 is a stunted cylinder and models repulsion since $\hat{\gamma}_1 < 1$, while $c(\cdot, r_2, t_2)$ is elongated and B_2 models aggregation, since $\hat{\gamma}_2 > 1$. Hence, when standing in some point $(x_1, y_1, z_1) \in X$ it is less likely to observe a z-coordinate if the corresponding xy-coordinates are similar to (x_1, y_1). Specifically, if (x_1, y_1) and (x_2, y_2) lies within distance 20 µm for `L3` and 24.25 µm for `L5`, it is less likely to observe a z-coordinate z_2 (associated to (x_2, y_2)) with $|z_1 - z_2|$ less than 11.5 µm for `L3` and 15.5 µm for `L5`. Analogously, given that (x_1, y_1) and (x_2, y_2) lies within distance 11 µm for `L3` and 14.75 µm `L5`, it is more likely to observe z_2 if $|z_1 - z_2|$ is in the interval from 11.5 µm to 35.5 µm for `L3` or from 15.5 µm to 37.25 µm for `L5`.

Finally, note that simulations from each of models 1–5 can straightforwardly be obtained using a Metropolis-Hastings algorithm for a fixed number of points and given a realisation of the xy-coordinates. Specifically, we used (Algorithm 7.1 in Møller and Waagepetersen, 2004) but with a systematic updating scheme cycling over the point indexes 1 to n, using a uniform proposal for a new point in W_z and a Hastings ratio calculated from the full conditional (C.3). We successively updated each point 100 times under the systematic updating scheme, corresponding to 63400 and 54800 point updates for `L3` and `L5`, respectively.

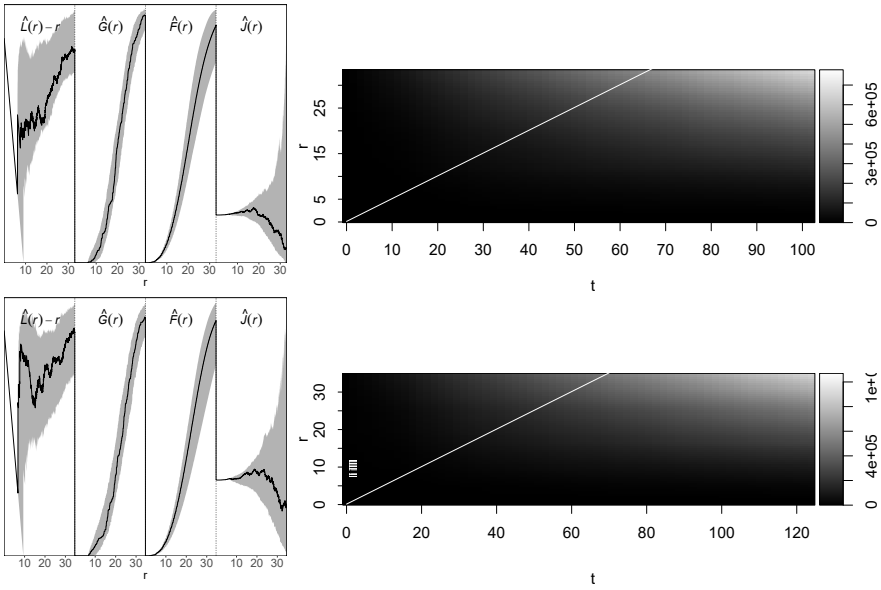

Fig. C.7: Results of the GERL envelope procedure under the fitted model 5 based on a concatenation of the empirical *L*-, *G*-, *F*-, *J*-, and cylindrical *K*-functions. Left: concatenation of the one-dimensional empirical functional summaries for the data (solid line) together with 95% envelopes (grey region); for ease of visualisation, the functions have been scaled. Right: empirical cylindrical *K*-function (grey scale) where shaded vertical/horizontal lines indicate that the function falls above/below the 95% envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the dataset `L3`. Bottom: results for the dataset `L5`.

Acknowledgements. This work was supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 'Statistics for point processes in space and beyond', and by the 'Centre for Stochastic Geometry and Advanced Bioimaging', funded by grant 8721 from the Villum Foundation. We are thankful to Ali H. Rafati for collecting the data analysed in this paper and to Jens R. Nyengaard and Ninna Vihrs for helpful comments.

References

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24:179–195.

Biscio, C. A. N. and Lavancier, F. (2016). Quantifying repulsiveness of determinantal point processes. *Bernoulli*, 22:2001–2028.

Buxhoeveden, D. P. and Casanova, M. F. (2002). The minicolumn hypothesis in neuroscience. *Brain*, 125:935–951.

Casanova, M. F. (2007). Schizophrenia seen as a deficit in the modulation of cortical minicolumns by monoaminergic systems. *International Review of Psychiatry*, 19:361–372.

Casanova, M. F., van Kooten, I. A. J., Switala, A. E., van Engeland, H., Heinsen, H., Steinbusch, H. W. M., Hof, P. R., Trippe, J., Stone, J., and Schmitz, C. (2006). Minicolumnar abnormalities in autism. *Acta Neuropathologica (Berl)*, 112:287–303.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer-Verlag, New York, second edition.

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Diggle, P. J. and Gratton, R. J. (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 46:193–227.

Esiri, M. M. and Chance, S. A. (2006). Vulnerability to Alzheimer's pathology in neocortex: the roles of plasticity and columnar organization. *Journal of Alzheimer's Disease*, 9:79–89.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101:1502–1512.

Guan, Y. (2009). A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes. *Statistics and Its Interface*, 2:91–99.

Lavancier, F., Møller, J., and Rubak, E. (2015). Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77:853–877.

Lavancier, F., Poinas, A., and Waagepetersen, R. P. (2018). Adaptive estimating function inference for non-stationary determinantal point processes. Available on arXiv:1806.06231.

van Lieshout, M. N. M. and Baddeley, A. (1996). A nonparametric measure of spatial interaction in point pattersn. *Statistica Neerlandica*, 50:344–361.

Møller, J., Christensen, H. S., Cuevas-Pacheco, F., and Christoffersen, A. D. (2019). Structured space-sphere point processes and K-functions. *Methodology and Computing in Applied Probability*. Available at https://doi.org/10.1007/s11009-019-09712-w.

Møller, J. and Christoffersen, A. D. (2018). Pair correlation functions and limiting distributions of iterated cluster point processes. *Journal of Applied Probability*, 55:789–809.

Møller, J., Safavimanesh, F., and Rasmussen, J. G. (2016). The cylindrical K-function and Poisson line cluster point process. *Biometrika*, 103:937–954.

Møller, J. and Torrisi, G. L. (2005). Generalised shot noise Cox processes. *Advances in Applied Probability*, 37:48–74.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, Boca Raton, Florida.

Mrkvička, T., Myllymäki, M., and Hahn, U. (2017). Multiple Monte Carlo testing, with applications in spatial point processes. *Statistics and Computing*, 27:1239–1255.

Mrkvička, T., Soubeyrand, S., Myllymäki, M., Grabarnik, P., and Hahn, U. (2016). Monte Carlo testing in spatial statistics, with applications to spatial residuals. *Spatial Statistics*, 18:40–53.

Mrkvička, T., Hahn, U., and Myllymäki, M. (2018). A one-way ANOVA test for functional data with graphical interpretation. Available on arXiv:1612.03608.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:381–404.

Ogata, Y. and Katsura, K. (1991). Maximum likelihood estimates of the fractal dimension for random spatial patterns. *Biometrika*, 78:463–474.

Prokešová, M., Dvořák, J., and Jensen, E. B. V. (2016). Two-step estimation procedures for inhomogeneous shot-noise Cox processes. *Annals of the Institute of Statistical Mathematics*, 69:1–30.

Rafati, A. H., Safavimanesh, F., Dorph-Petersen, K. A., Rasmussen, J. G., Møller, J., and Nyengaard, J. R. (2016). Detection and spatial characterization of minicolumnarity in the human cerebral cortex. *Journal of Microscopy*, 261:115–126.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13:255–266.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, 36:18–25.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252–258.

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-487-4 AALBORG UNIVERSITY PRESS